生命科学名著

Mc Graw Hill Education

〔美〕 Robert F. Weaver 著

郑用琏 张富春
徐启江 岳 兵 主译

分子生物学

（原书第四版）

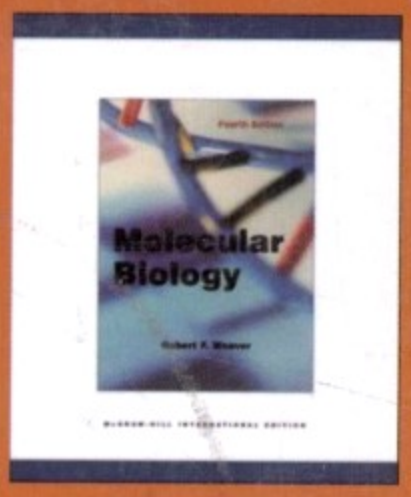
Molecular Biology

(4th Edition)

科学出版社
www.sciencep.com

Mc Graw Hill

(Q-2397.0101)



分子生物学是生命科学发展过程中诞生的一门实验性极强的新兴学科。美国著名分子生物学家Robert F. Weaver遵循这一学科发展的特点，于1999年出版了*Molecular Biology*一书的第一版。全书以原始研究论文为基础，通过对实验的设计、对结果的分析而逐步展开对分子生物学理论的讲述，文字通俗流畅，叙述由浅入深。随着学科的迅速发展，几经修订再版的*Molecular Biology*第四版共有导论，分子生物学方法，原核生物、真核生物转录，转录后加工，翻译，DNA复制、重组和转座，以及基因组学等8部分24章，书后还附有词汇表。每一章节都以提出科学问题、展开研究过程开始，以提供思考习题、推荐阅读文献结束，理论讲述逻辑严密，实验过程提炼清晰，特色鲜明、内容详尽，图文并茂，易读易记。本书是生命科学相关专业的本科生、研究生及从事该方面科研、教学人员不可多得的一本参考用书。

生物分社
联系电话：010-64012501
http://www.lifescience.com.cn
e-mail:lifescience@mail.sciencep.com

销售分类建议：生物/分子生物学

http://www.mheducation.com

ISBN 978-7-03-026194-6

定 价：145.00 元

生命科学名著

分子生物学

（原书第四版）

Molecular Biology（4th Edition）

〔美〕Robert F. Weaver 著

郑用琏 张富春 徐启江 岳 兵 主译

Q7-43
W51

科学出版社

北京

]字：01-2008-0514 号

内 容 简 介

分子生物学是生命科学发展过程中诞生的一门实验性极强的新兴学科。美国著名分子生物学家 Robert F. Weaver 遵循这一学科发展的特点，于 1999 年出版了第一版 *Molecular Biology*。全书以原始研究论文为基础，通过对实验的设计、对结果的分析而逐步展开对分子生物学理论的讲述，文字通俗流畅。随着学科的迅速发展，几经修订再版的 *Molecular Biology* 第四版共有导论，分子生物学方法，原核生物的转录，真核生物的转录，转录后加工，翻译，DNA 复制、重组和转座，以及基因组 8 部分 24 章，书后还附有分子生物学专业词汇表。每一章节都以提出科学问题、展开研究过程开始，以提供思考习题、推荐阅读文献结束，理论讲述逻辑严密，实验过程提炼清晰，特色鲜明，内容详尽，图文并茂，易读易记。

本书是生命科学相关专业的研究生以及从事该方面科研、教学人员不可多得的一本优秀参考书。

Robert F. Weaver
Molecular Biology, Fourth edition
ISBN: 978-0-07-299524-4
Copyright ©2008 by The McGraw-Hill Companies, Inc.

All Rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including without limitation photocopying, recording, taping, or any database, information or retrieval system, without the prior written permission of the publisher.

This authorized Chinese translation edition is jointly published by McGraw-Hill Education (Asia) and Science Press. This edition is authorized for sale in the People's Republic of China only, excluding Hong Kong, Macao SAR and Taiwan.

Copyright ©2009 by McGraw-Hill Education (Asia), a division of the Singapore Branch of The McGraw-Hill Companies, Inc. and Science Press.

版权所有。未经出版人事先书面许可，对本出版物的任何部分不得以任何方式或途径复制或传播，包括但不限于复印、录制、录音，或通过任何数据库、信息或可检索的系统。

本授权中文简体字翻译版由麦格劳-希尔（亚洲）教育出版公司和科学出版社合作出版。此版本经授权仅限在中华人民共和国境内（不包括香港特别行政区、澳门特别行政区和台湾）销售。

版权©2009 由麦格劳-希尔（亚洲）教育出版公司与科学出版社所有。

本书封面贴有 McGraw-Hill 公司防伪标签，无标签者不得销售。

图书在版编目(CIP)数据

分子生物学/（美）Weaver，R. F. 著，郑用琏等译. —4 版. —北京：科学出版社，2009
（生命科学名著系列）
ISBN 978-7-03-026194-6

Ⅰ. 分… Ⅱ. ①W…②郑… Ⅲ. 分子生物学-高等学校-教材 Ⅳ. Q7

中国版本图书馆 CIP 数据核字（2009）第 225661 号

责任编辑：李 悦 席 慧/责任校对：钟 洋
责任印制：钱玉芬/封面设计：陈 敬

科学出版社 出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com

新蕾印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2010 年 1 月第 一 版 开本：787×1092 1/16
2010 年 1 月第一次印刷 印张：52 1/2
印数：1—4 000 字数：1 223 480

定价：145.00 元（含光盘）

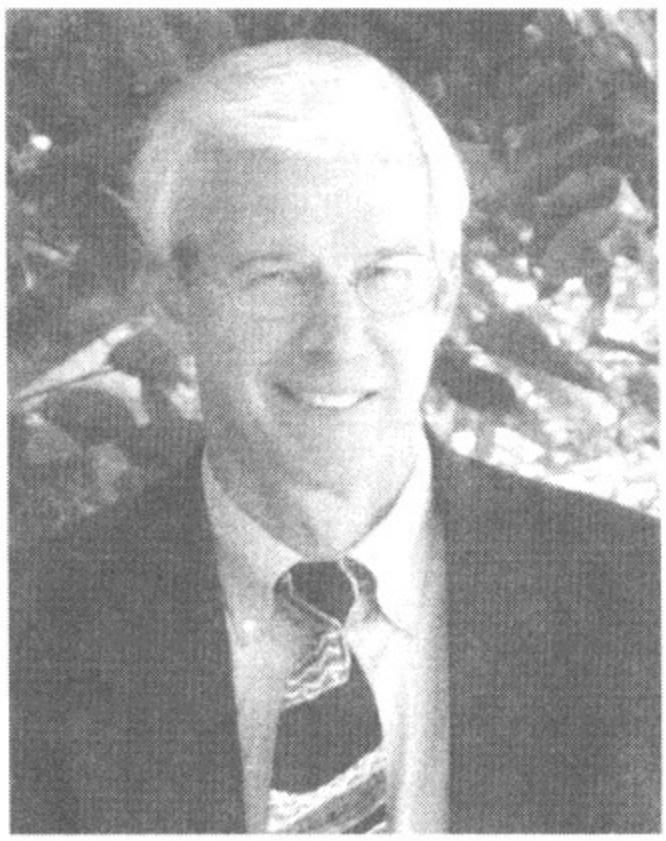
（如有印装质量问题，我社负责调换〈路通〉）

分子生物学（原书第四版）
译校人员名单

（按姓氏汉语拼音排序）

陈　璇　高友军　兰海燕　李玉花
刘智捷　马　纪　马正海　肖海林
徐启江　许志茹　岳　兵　张富春
张绍鹏　张姿丽　赵衍鑫　郑用琏
周　波

作者简介



Robert F. Weaver 出生于美国堪萨斯州的首府托皮卡市，在弗吉尼亚州的阿灵顿地区长大。1964 年在俄亥俄州的乌斯勒学院获得化学学士学位。1969 年在杜克大学获得生物化学专业博士学位，此后他在加州大学旧金山分校从事了两年的博士后研究工作，师从 William J. Rutter 教授研究真核生物 RNA 聚合酶的结构。

1971 年他受聘于堪萨斯大学，任生物化学助理教授，后晋升为副教授，并于 1981 年任教授。自 1984 年以来，Robert F. Weaver 一直担任生物化学系的系主任，1995 年开始担任文理学院副院长。

文理学院管辖 14 个不同的系和研究中心，Weaver教授分管科学教育系和数学系。作为一位分子生物学教授，他主讲分子生物学概论和肿瘤分子生物学两门课程，并指导本科生和研究生在实验室进行感染鳞翅目幼虫的杆状病毒分子生物学方面研究工作。

Weaver 教授发表了多篇科学论文，他的研究受到美国国立卫生研究院、美国国家科学基金和美国癌症协会的资助，并且他还与同事一起合著了两本遗传学教科书，为《美国国家地理杂志》撰写过两篇分子生物学方面的文章。作为美国癌症协会的研究人员，他在欧洲的两个实验室（瑞士的苏黎世和英国的牛津大学）分别进行了一年的访问研究。

译 者 序

分子生物学是在分子水平研究基因的结构和功能的科学，是生命科学发展过程中诞生的一门新兴学科，特别是随着生物技术发展的日新月异，分子生物学的理论也得以迅速发展与不断完善。分子生物学作为一门实验性很强的学科，它的每一个结论、每一个概念都源自研究者对大量科学实验结果的总结与升华。Robert F. Weaver 博士的 *Molecular Biology* 一书从始至终贯穿了分子生物学这一学科的发展规律。正如作者在序言中开宗明义地表明了他的撰写初衷："I really wanted a textbook that presented the concepts of molecular biology, along with the experiments that led to those concepts."（译文：我希望这样的教科书，它能够清晰地陈述分子生物学的概念以及提出这些概念的实验。）也正是基于这一不同于其他分子生物学教材的特点，中国科学院推荐这一名著作为中国科学院研究生教学用书，科学出版社于 2000 年以影印版的方式第一次出版发行了 *Molecular Biology* 一书。在科学出版社编辑的倡导下，译者以导读版的形式将 Weaver 教授 2005 年撰著的 *Molecular Biology* 第三版编译出版，时隔三年，*Molecular Biology* 的第四版又问世了。

第四版 *Molecular Biology* 不仅对所有章节的内容进行了更新，补充了学科发展的最新成果与进展，而且对全书的图示与图解进行了更加清晰、更为明了的设计与阐释。全书最大特色是，对每一个分子生物学结论的介绍都是通过对实验的设计、对结果的分析而逐步展开的。全书以原始研究论文为基础，文字通俗流畅，叙述由浅入深，每一章节都以提出科学问题、展开研究过程开始，以提供思考习题、推荐阅读文献结束，很适合学生的阅读与理解，更利于知识的巩固与提高。凡是认真学习过 *Molecular Biology* 英文教材的学生，都普遍感到从中受益的不仅是准确地掌握了分子生物学的基本理论和学科前沿，更为重要的是得到开展分子生物学理论研究的方法、思维、逻辑与分析的启迪，以及创新能力的提升。

正因为 *Molecular Biology* 第四版的理论讲述逻辑严密，实验过程提炼清晰，结论归纳严谨准确，译者唯恐有失作者高超的写作功底，精彩的逻辑推理……因此，我们不仅组织了华中农业大学生命科学技术学院、新疆大学生命科学院和东北林业大学生命科学院一直以 Robert F. Weaver 的 *Molecular Biology* 一书为教学参考书、长期从事分子生物学教学的教师以及部分博士研究生参加翻译工作，而且所有译者都有编译第三版时通读原著全文并精读部分章节的经历与感悟。特别是当 *Molecular Biology* 第四版初译稿完成后，我们又将初稿分发给在读研究生，征求意见，查找疏漏，更臻完善。

然而由于译者水平有限，时间仓促，不乏错误之处，恳望读者建议指正。

郑用琏　张富春　徐启江　岳　兵

2009 年 4 月

序　言

本书是分子生物学的介绍性教程。但什么是分子生物学？这个难以琢磨的定义依赖于谁在下定义。在本书中，我认为分子生物学是一门在分子水平研究基因及其活性的学科。

当我还是大学生和研究生的时候，每当老师强调实验的策略、强调得出科学结论的实验结果与过程而非结论本身时，我就会因探求科学奥秘的过程而异常兴奋，激发我特别努力地学习。因此当我在 1972 年开始讲授分子生物学导论课程时，就采用了这种讲授策略，并一直延用至今。我发现学生也像我当年一样反应积极。

然而，这种方法存在的一个问题是没有一本教科书能像我想做的那样，强调实验数据的重要性，因此我尝试通过指定阅读文献来代替教科书。虽然这一方式完全适用于高级课程的讲授，但对分子生物学的初级课程来说，效率却相对较低，并且也不实用。为了改进这一过程，我把要讲解的数据以手绘卡通图片的形式来充实推荐阅读的文献，将文章中的图和表做成透明胶片，但我还是希望有一本教科书在给出分子生物学概念的同时，能介绍得出这些概念的实验，并且清楚地向学生解释实验和概念之间的关系。然而，最终我意识到获得这种书的最好办法是自己来编写。我曾经成功地参与编写过一本有关遗传学导论的教科书，在那本书中我尽量采取了从实验入手的思路进行编写。这一经历也给了我勇气去尝试自己写一本书，并将这一思路作为本书的一个创新性的探索。

章节安排

本书开始的 4 章，对大多数学生来讲是个回顾。第 1 章是遗传学发展简史，第 2 章讨论 DNA 的结构和化学性质，第 3 章是基因表达概述，第 4 章介绍基因克隆的具体细节。这些内容是大多数分子生物学学生在遗传学导论课程中已经学习过的，但分子生物学学生仍然需要掌握这些概念，也需要知识更新。我在课堂不专门讲授这几章，而是建议学生如果需要可以参考。

第 5 章介绍了几种分子生物学常用的技术。要想在一章中涵盖本书所描述的所有技术是不可能的，所以我试图概括最普通的，或少数情况下书中其他地方没有提及的最有价值的技术。我在讲课时，没有按书的内容讲第 5 章，而是在以后的章节中当首次遇到某个技术时才让学生查阅，这样做可使学生避免对接二连三的技术感到厌倦。我也意识到在这些技术中包含着一些十分复杂的概念，学生只有在获得更多分子生物学的实践后才能有更深刻的理解和感悟。

第 6～9 章介绍原核生物的转录。第 6 章介绍基本的转录装置，包括启动子、终止子、RNA 聚合酶，展示转录是如何启动、延伸和终止的。第 7 章介绍在三种不同操纵子中转录的调控。第 8 章介绍细菌和噬菌体如何在同一个时间控制多个基因的转录，如它们常常通过提供不同的 σ 因子来控制这一转录过程。第 9 章讨论细菌的 DNA 结合蛋白（多数为螺旋-转角-螺旋蛋白）与其靶 DNA 的相互作用。

第 10～13 章描述真核生物的转录控制。其中，第 10 章涉及 3 种真核生物 RNA 聚合酶及其所识别的启动子。第 11 章介绍与 3 种 RNA 聚合酶协作的通用转录因子，指出 TATA 盒-结合蛋白的一致性规律，它们参与所有 3 种聚合酶的转录。第 12 章解释了基因特异性转录因子或激活因子的功能，还介绍了几种代表性激活因子的结构，显示它们如何与其靶 DNA 相互作用。第 13 章描述真核生物染色质的结构，显示激活物如何与组蛋白相互作用激活或抑制转录。

第 14～16 章介绍真核生物的转录后加工。其中，第 14 章是关于 RNA 剪接的；第 15 章描述加帽和多聚腺苷酸化；第 16 章介绍一个令人着迷的“其他转录后事件”的综观，包括 rRNA 和 tRNA 加工、反式剪接和 RNA 编辑。第 16 章也讨论 3 种基因表达的转录后调控：①RNA 干涉；②调节 mRNA 的稳定性（以转铁蛋白受体基因为引发例子）；③micro RNA 实施的调控。

第 17～19 章描述在细菌和真核生物中的翻译过程。其中，第 17 章涉及翻译的起始，包括翻译起始的控制；第 18 章讲述多肽是如何延伸的，重点是在细菌中的延伸；第 19 章介绍翻译过程中核糖体和 tRNA 这两个关键因子的结构和功能。

第 20～23 章阐述 DNA 复制、重组和转座的机制。其中，第 20 章介绍 DNA 复制和修复的基本机制，以及复制过程的有关蛋白质（包括 DNA 聚合酶等）。第 21 章详述在细菌和真核生物中 DNA 复制的起始、延伸和终止步骤。第 22 章和第 23 章描述细胞中自然发生的 DNA 重排。第 22 章讨论同源重组，第 23 章涉及转座。

第 24 章介绍基因组学和蛋白质组学的概念。该章以一个古老的基因图位克隆故事开始，故事涉及利用人类基因组信息（或其他基因组），对 Huntington 疾病基因采用较为漫长、枯燥而且相对简单的图位克隆策略进行定位的过程。

第四版新增内容

第三版出版以来，分子生物学最突出的发展是对非编码 RNA 重要性的认识，如 microRNA（miRNA）调控真核基因的表达。我相应地在第 16 章中加入了新的一节，展示 miRNA 如何像 siRNA 一样引起目标 mRNA 的降解，或抑制目标 mRNA 的翻译但不破坏 mRNA。

这一版还介绍了核糖体开关，结合到小分子上的 mRNA 片段通过改变 RNA 构象做出响应，从而控制翻译。在第 7 章中会看到一个例子，小分子结合到新生 mRNA 的核糖体开关上，迫使其形成终止子，中断该 mRNA 的合成。在第 17 章还会看到另一个例子，其中小分子结合到 mRNA 的核糖体开关上引起 mRNA 构象的改变，进而隐藏核糖体结合位点，阻断翻译过程。

第四版的所有各章除导论部分外都进行了更新，补充了新信息，以下是几个重点。

第 10 章：与 DNA 结合的 RNA 聚合酶 II 的 X 射线晶体学揭示了聚合酶分离新生 RNA 与其 DNA 模板的作用。12 亚基的 RNA 聚合酶 II 首次被结晶出来。这一完整的晶体结构显示存在一个能让 DNA 单链分子进出的较深的裂缝。这表明 DNA 必须解链，以使模板链能进入活性位点。

第 11 章：对 TFIIB 聚合酶 II 复合物的结构学研究显示，TFIIB 通过其 C 端结构域在 TATA 框与 TBP 结合，而聚合酶 II 则是通过其 N 端。这一搭桥作用保证了聚合酶活性中心能粗略地定位在 TATA 框下游 25bp 处。

第 13 章：虽然我们还不很清楚染色质重构的细微末节，但现已知道被催化的核小体重构过程涉及与核小体 DNA 结合的核心组蛋白的不同构象的形成。这与未催化的暴露在核小体中，或核小体简单地沿一段 DNA 滑动相反。我们对 30nm 纤丝的结构也有了新的认识，四聚核小体的晶体结构由双摞核小体组成，两摞之间以 Z 字形 DNA 连接。这种排列与螺线管模型及其他关于 30nm 纤丝的模型都不同，但与交联体、双起始螺旋一致，即核小体的两摞在左手双螺旋上可以相互缠绕。

第 14 章：高等真核生物中选择性剪切是一个普遍现象，它提供了一种由相同基因产生更多蛋白质产物的方式，以及细胞内基因表达调控的方式。选择性剪切的控制由结合到剪切位点和分支位点的剪切因子实施，也由那些与外显子剪切增强子（ESE）、外显子剪切沉默子（ESS）和内含子沉默元件相互作用的蛋白质实施。

第 16 章：除了在 miRNA 方面的新内容以外，这一章还包括 RNAi 机制的新信息。具体来讲，作为 RNA 切割酶（slicer）切割目标 mRNA 的蛋白质，现已鉴定为 Argonaute2，是一种最小的 RISC，只由 Argonaute 蛋白 Ago2 和 siRNA 组成，关于 RISC 组装的新信息也进行了介绍。

第 17 章：除了关于由核糖体开关控制的原核生物翻译起始的新内容以外，还给出了 microRNA对真核生物翻译起始调控的模型。

第 18 章：该章讨论了稀有硒代半胱氨酸和吡咯赖氨酸分别响应终止密码子 UGA 和 UAG 而掺入到延伸的多肽中的过程，也讨论了一种称作激发因子的原核蛋白，它与核糖体联合，俘获从核糖体的出口通道出现的新生多肽，以此保护新生多肽的疏水区直到出现合适的能与之结合的搭档为止。

第 19 章：该章给出了 P 位点 tRNA 2′-羟基在转肽作用中的化学参与。还讨论了 *E. coli* 核糖体的第一个晶体结构，揭示了 30S 颗粒头部的扭转与移位的可能关系。

第 20 章：该章介绍了两种新的碱基切除修复（DNA 中 8-羟基鸟嘌呤的修复）机制。

第 22 章：讨论 Spo11 在引发并共价结合到双链 DNA 之后，从 DSB 与 12～27nt 寡核苷酸复合物中的释放。Spo11 的切割看起来是非对称的，提示了随后的 Holliday 联接体形成的机制。

第 24 章：基因组学、蛋白质组学及生物信息学领域正以爆炸性的速度增长。相应地，第 24 章在第四版中是更新最彻底的一章，名称也改变了，以便包括生物信息学的内容和反映这个新领域的重要性。该章更新的一些重点包括以下内容：①全基因组转录图谱，高分辨率的转录物作图，染色体的具体位点。这一研究已揭示出非编码 RNA 的丰富性，包括真核细胞中的非多聚腺苷酸化 RNA。②RNAi 分析，RNAi 用于在基因组范围的基础上抑制基因表达及抑制效果的分析。③组织特异性表达概况，这种表达状况可通过检测由外源 miRNA 的干扰而降低表达水平的 mRNA 谱而获得。④国际 HapMap 协会已经发表了单倍型图谱，包括 100 万个人类 SNP，这些 SNP 是通过对 4 个有明显地域差异的人类种群进行 269 个 DNA 样本的基因分型后获得的。⑤科学家利用生物信息学技术已经发现了 4 种哺乳动物（包括人类）启动子区和 3′-UTR 区的高度保守的序列。启动子区的基序可能为转录因子提供了结合位点。3′-UTR 区的多数基序可能为 miRNA 提供了结合位点。⑥已完成的人类基因组草图是一个伟大的成就，讨论了这一成就的重要性。讨论了人类 X 染色体的特征，该染色体与人类疾病有关。

补充说明

McGraw-Hill 展示中心

随时、随地、任意方式建立你的教学材料！ ARIS 展示中心是一个在线数字式图书馆，资源有图片、艺术品、幻灯片、模拟动画以及其他媒体类型的材料，可用于创建个性化的讲座、可视化增强的考试或测验、必修课程的网页或有吸引力的印刷辅助材料。

再简单不过了！ 在教科书的 ARIS 网站中，从“教师”窗口进入，展示中心的动态引擎将为你提供按学科、课程、教材章节、资源类型、关键词的搜索功能。只需简单地浏览、选择，然后下载所需文件便可建立漂亮的课程材料。所有材料都归 McGraw-Hill 高等教育出版社版权所有，但教师可以用于课堂教学。

教师的试题和 CD-ROM 资源*

ISBN-13：978-0-07-299527-5

* 采用本书作为教材的教师可向 McGraw-Hill 公司北京代表处联系索取教学课件资料。详情请见本书所附“教师反馈表”。

ISBN-10：0-07-299527-0

这个交叉平台 CD 以计算机化的试题库为特征，利用试题软件迅速生成个性化的试卷。用户友好程序可以按题目、格式、难度水平、编辑现有题目、增加新题等方式检索，将相同的试题混合生成各种题型。

教材专用网站

下列网站是英文原版教材的专用网站，提供有权使用的数字成像文件、模拟动画、在线 PowerPoint 讲座、各章复习题的答案以及网页链接。

www. mhhe. com/weaver4

翻译　马　纪　校译　张富春　郑用琏

致　谢

在编写这本参考书的过程中，许多编辑和审稿人给了我很大的帮助。他们的建议大大增加了这本书的准确性和可读性，但他们没有义务对书中仍然存在的任何错误和模糊之处负责，对此，我应负全部的责任。我衷心感谢以下人员的帮助。

第四版审稿人：
Dr. David Asch
Youngstown State
University

Christine E. Bezotte
Elmira College
Mark Bolyard
Southern Illinois University, Edwardsville

Diane Caporale
University of Wisconsin,
Stevens Point

Jianguo Chen
Claflin University
Chi-Lien Cheng
Department of Biological Sciences,
University of Iowa
Mary Ellard-Ivey
Pacific Lutheran University Olukemi Fadayomi
Ferris State University, Big Rapids, Michigan

Charles Giardina
University of Connecticut Eli V. Hestermann
Furman University

Dr. Dorothy Hutter
Monmouth University
Cheryl Ingram-Smith
Clemson University

Dr. Cynthia Keler
Delaware Valley College
Jack Kennell
Saint Louis University

Charles Ii. Mallory
College of Arts and
Sciences, University of
Miami

Jon L. Milhon
Azusa Pacific University

Hao Nguyen
California State University, Sacramento

Thomas Peterson
Iowa State University
Ed Stellwag
East Carolina University

Katherine M. Walstrom
New College of Florida
Cornelius A. Watson
Roosevelt University
Fadi Zaher
Gateway Technical
College

第三版审稿人：

David Asch
Youngstown State
University

Gerard Barcak
University of Maryland
School of Medicine

Bonnie Baxter
Hobart Be William Smith Colleges
André Bédard
McMaster University' Felix

Breden
Simon Fraser University
Laura Bull
UCSF Liver *Center*
Laboratory
James Ellis
Developmental Biology
Program, Hospital for Sick Children, Toronto, Ontario
Robert Helling
The University of
Michigan

David Hinkle
University of Rochester

Robert Leamnson
University of Massachusetts at Dartmouth

David Mullin
Tulane University
Marie Pizzorno
Bucknell University
Michael Reagan
College *of* St. Benedict/
St. John's University
Rodney Scott
Wheaton College

第二版审稿人：

Mark Bolyard
Southern Illinois University M. Suzanne Bradshaw
University of Cincinnati
Anne Britt -
University of California,
Davis

Robert Brunner
University of California,
Berkeley
Caroline J. Decker
Washington State University

Jeffery DeJong
University of Texas, Dallas Stephen J. D'Surney
University of Mississippi
John S. Graham
Bowling Green State
University
Ann Grens
Indiana University
Ulla M. Hansen
Boston University
Laszlo Hanzely
Northern Illinois University Robert B. Helling
University of Michigan

Martinez J. Hewlett
University of Arizona

David C. Hinkle
University of Rochester
Barbara C. Hoopes
Colgate University

Richard B. Imberski
University of Maryland
Cheryl Ingram-Smith
Pennsylvania State
Universi ty
Alan Kelly
University of Oregon

Robert N. Leamnson
University of
Massachusetts, Dartmouth
Karen A. Malatesta
Princeton University
Robert P. Metzger
San Diego State University

David A. Mullin
Tulane University
Brian K. Murray
Brigham Young University

Michael A. Palladino
Monmouth University
James G. Patton
Vanderbilt University
Martha Peterson
University of Kentucky
Marie Pizzorno
Bucknell University

Florence Schmieg
University of Delaware

Zhaomin Yang
Auburn University

第一版审稿人：

Kevin L. Anderson
Mississippi State University

Rodney I'. Anderson
Ohio Northern University

Prakash H. Bhuta
Eastern Washington
University

Dennis Bogyo
Valdosta State University

Richard Crawford
Trinity College

Christopher /\. Cullis
Case Western Reserve
University

Beth De Stasio
Lawrence University

R. Paul Evans
Brigham Young University
Edward R. Fliss
Missouri Baptist College

Michael A. Goldman
San Francisco State
University

Robert Gregerson
Lyon College

Eileen Gregory
Rollins College

Barbara A. Hamkalo
University of California, Irvine

Mark L. Hammond
Campbell University

Terry L. Helser
State University of New York,
Oneonta

Carolyn Herman
Southwestern College

Andrew S. Hopkins
Alverno College

Carolyn Jones
Vincennes University

Teh-Hui Kao
Pennsylvania State
University
Mary Evelyn B. Kelley
Wayne State University

Harry van Keulen
Cleveland State University

Leo Kretzner
University of South Dakota

Charles J. Kunert
Concordia University

Robert N. Leamnson
University of
Massachusetts, Dartmouth

James D. Liberatos
Louisiana Tech University

Cran Lucas
Louisiana State University

James J. McGivern
Gannon University

James E. Miller
Delaware Valley College

Robert V. Miller
Oklahoma State University

George S. Mourad
Indiana University-Purdue Univer-
sity

David A. Mullin
Tulane University
James R. Pierce
Texas A8cM University,

Kingsville
Joel B. Piperberg
Millersville University

John E. Rebers
Northern Michigan
University

Florence Schmieg
University of Delaware

Brian It. Shmaefsky
Kingwood College

Paul Keith Small
Eureka College

David J. Stanton
Saginaw Valley State
University

Francis X. Steiner
Hillsdale College

Amy Cheng Vollmer
Swarthmore College

Dan Weeks
University of Iowa

David B. Wing
New Mexico Institute of Mining Sc
Technology

分子生物学实验技术目录

实验技术	章	页码
转化	2	10
超速离心法	2	11
杂交	2	22
指纹(蛋白质)	3	39
质粒载体基因克隆	4	45
筛选	4	46
影印培养法	4	46
λ噬菌体载体基因克隆	4	49
噬菌斑杂交	4	50
M13 噬菌体载体基因克隆	4	51
黏粒载体基因克隆	4	51
噬菌粒载体基因克隆	4	52
cDNA 克隆	4	53
寡核苷酸探针设计	4	53
cDNA 末端快速扩增(RACE)	4	55
菌落杂交	4	55
切口平移	4	55
反转录 PCR (RT-PCR)	4	56
聚合酶链式反应(PCR)	4	56
实时定量 PCR	4	59
表达载体	4	60
亲和柱层析法	4	61
Ti 载体基因克隆	4	64
杆状病毒表达载体	4	64
DMS 足迹法	5	100
SELEX (指数级富集配体系统进化技术)	5	101
功能性 SELEX	5	101
敲除小鼠	5	101
凝胶电泳(DNA)	5	69
SDS-PAGE (蛋白质)	5	71
脉冲场凝胶电泳(PFGE)	5	71
凝胶电泳(蛋白质)	5	71
等电聚焦	5	72
双向凝胶电泳	5	72
离子交换层析	5	73
凝胶过滤层析	5	73
放射性自显影	5	74
磷屏成像	5	75
液体闪烁计数	5	76
Southern 印迹	5	77
DNA 分型	5	78
DNA 指纹	5	79
荧光原位杂交(FISH)	5	80
原位杂交	5	80
免疫印迹(Western 印迹)	5	81
DNA 测序(Sanger 法)	5	82
限制性图谱	5	83
自动化 DNA 测序	5	83
定点诱变	5	87
Northern 印迹	5	89
S1 图谱定位	5	90
末端补平法	5	91
RNase 作图 (RNase 保护分析)	5	92
引物延伸	5	92
截断转录	5	93
无 G 盒转录	5	93
斑点印迹	5	94
报告基因转录分析	5	94
连缀转录	5	94
滤膜结合分析 (DNA-protein 相互作用)	5	97
免疫沉淀法	5	97
DNase 足迹法	5	98
凝胶阻滞分析	5	98
亲和标记	6	126
DNA-蛋白质交联	6	129
电泳检测 DNA 弯曲	7	160
X 射线晶体学	9	198
表位附加法	10	221
接头分区突变	10	234

实验技术	章	页码
活性胶分析	13	331
染色质免疫沉淀(ChIP)	13	337
R 环	14	350
RNA-RNA 交联(与 4-硫尿嘧啶)	14	360
RNA-RNA 交联 (与补骨脂素)	14	361
寡核苷酸指导的 RNA 降解	14	364
合成致死筛选法	14	375
酵母双杂交实验	14	376
酵母双杂交筛选	14	376
Far Western 印迹	15	419
脉冲示踪标记	16	445
RNA 干涉(RNAi)	16	451
RNA 解旋酶	17	490
趾纹分析	17	494
停流装置动力学实验	18	528
通过蛋白质微测序设计探针	18	540
等位基因特异性 RNAi	18	545
羟基自由基探针	18	547
CsCl 密度梯度超速离心	20	580
DNA 解旋酶分析	20	596
拓扑异构酶分析	20	599
蛋白质足迹法	21	639
图位克隆技术	24	704
外显子捕获	24	705
限制性片段长度多态性(RFLP)	24	705
酵母人工染色体基因克隆	24	714
细菌人工染色体基因克隆	24	715
可变数串联重复序列(VNTR)	24	716
微卫星序列	24	716
序列标签位点(STS)	24	716
表达序列标签(EST)	24	718
辐射杂种作图法	24	718
鸟枪法测序	24	718
DNA 微阵列	24	729
DNA 芯片	24	730
基因表达系列分析(SAGE)	24	733
噬菌体展示	24	746

简要目录

译者序
序言
致谢
分子生物学实验技术目录

第Ⅰ部分　导论

1　分子生物学简史
2　基因的分子特性
3　基因功能简介

第Ⅱ部分　分子生物学方法

4　分子克隆方法
5　研究基因及基因活性的分子工具

第Ⅲ部分　原核生物的转录

6　细菌的转录装置
7　操纵子：细菌转录的精细调控
8　细菌转录机制的主要改变
9　细菌中 DNA-蛋白质的相互作用

第Ⅳ部分　真核生物的转录

10　真核生物的 RNA 聚合酶及其启动子
11　真核生物中的通用转录因子
12　真核生物的转录激活因子
13　染色质结构及其对基因转录的影响

第Ⅴ部分　转录后加工

14　mRNA 加工 I：剪接
15　mRNA 加工 II：加帽和多聚腺苷酸化
16　其他 RNA 的加工过程

第Ⅵ部分　翻译

17　蛋白质翻译机制 I：起始
18　蛋白质翻译机制 II：延伸与终止
19　核糖体和转运 RNA

第Ⅶ部分　DNA 复制、重组和转座

20　DNA 复制 I：基本机制与酶学
21　DNA 复制 II：详细机制
22　同源重组
23　转座

第Ⅷ部分　基因组

24　基因组学、蛋白质组学和生物信息学

分子生物学专业词汇表
参考文献
索引
教师反馈表

翻译　马　纪　校译　张富春

目　录

译者序
序言
致谢
分子生物学实验技术目录

第Ⅰ部分　导　论

第 1 章　分子生物学简史 …… 1
1.1　传递遗传学 …… 1
孟德尔的遗传定律 …… 1
遗传的染色体理论 …… 2
遗传重组和遗传图谱 …… 3
重组的物理学证据 …… 4
1.2　分子遗传学 …… 4
DNA 的发现 …… 4
基因和蛋白质之间的关系 …… 5
基因的活性 …… 5
1.3　生命的 3 个领域 …… 7
第 2 章　基因的分子特性 …… 10
2.1　遗传物质的特性 …… 10
细菌转化 …… 10
多聚核苷酸的化学本质 …… 13
2.2　DNA 结构 …… 15
实验背景 …… 15
双螺旋 …… 15
2.3　RNA 基因 …… 18
2.4　核酸的物理化学性质 …… 19
DNA 结构的多样性 …… 19
不同大小和形状的 DNA …… 22
第 3 章　基因功能简介 …… 25
3.1　储存信息 …… 25
基因表达的总体过程 …… 25
蛋白质结构 …… 26
蛋白质的功能 …… 29
信使 RNA 的发现 …… 31
转录 …… 31
3.2　复制 …… 38
3.3　突变 …… 39
镰状细胞贫血病 …… 39

第Ⅱ部分　分子生物学方法

第 4 章　分子克隆方法 …… 42
4.1　基因克隆 …… 42
限制性内切核酸酶的作用 …… 42
载体 …… 45
用特异性探针鉴别特定的克隆 …… 52
cDNA 克隆 …… 53
cDNA 末端快速扩增 …… 55
4.2　聚合酶链式反应 …… 56
标准 PCR …… 56
cDNA 克隆中反转录 PCR(RT-PCR)的应用 …… 56
实时定量 PCR …… 59
4.3　表达克隆基因的方法 …… 59
表达载体 …… 60
其他真核载体 …… 64
利用 Ti 质粒将基因导入植物中 …… 64
第 5 章　研究基因及基因活性的分子工具 …… 69
5.1　分子的分离 …… 69
凝胶电泳 …… 69
双向凝胶电泳 …… 72
离子交换层析 …… 73
凝胶过滤层析 …… 73
亲和层析 …… 73
5.2　标记性示踪物 …… 74
放射性自显影 …… 74
磷屏成像 …… 75
液体闪烁计数器 …… 76
非放射性示踪剂 …… 76
5.3　利用核酸杂交 …… 77
Southern 印迹:鉴定特异性 DNA 片段 …… 77
DNA 指纹技术和 DNA 分型 …… 78
DNA 指纹技术和 DNA 分型在法医学中的应用 …… 79

Western 印迹 …… 81
DNA 测序 …… 81
限制性图谱 …… 83
基于克隆基因的蛋白质工程:定点诱变 …… 87
5.4 转录物的图谱定位与定量 …… 89
Northern 印迹 …… 89
S1 图谱定位 …… 90
引物延伸 …… 92
截断转录和无 G 盒转录 …… 93
5.5 检测体内转录效率 …… 94
核连缀转录分析 …… 94
报告基因转录分析 …… 94
体内蛋白质积累的检测 …… 96
5.6 分析 DNA-蛋白质的相互作用 …… 97
过滤结合 …… 97
凝胶阻滞 …… 98
DNA 酶足迹 …… 98
DMS 足迹法和其他足迹法 …… 98
5.7 寻找与其他分子相互作用的 RNA 序列 …… 101
SELEX …… 101
功能性 SELEX …… 101
5.8 基因敲除 …… 101

第Ⅲ部分 原核生物的转录

第 6 章 细菌的转录装置 …… 107
6.1 RNA 聚合酶的结构 …… 107
σ 亚基是一种特异性因子 …… 108
6.2 启动子 …… 109
RNA 聚合酶与启动子的结合 …… 109
启动子结构 …… 111
6.3 转录起始 …… 112
σ 因子的功能 …… 113
σ 因子的结构 …… 120
α 亚基在 UP 元件识别中的作用 …… 123
6.4 延伸 …… 125
核心酶在延伸过程中的作用 …… 125
延伸复合体的结构 …… 130
6.5 转录的终止 …… 139
rho 非依赖型终止 …… 139
rho 依赖型终止 …… 143
第 7 章 操纵子:细菌转录的精细调控 … 148
7.1 *lac* 操纵子 …… 148
lac 操纵子的负调控 …… 149
操纵子的发现 …… 150
阻遏物与操纵基因间的相互作用 …… 153
阻遏机制 …… 153
lac 操纵子的正调控 …… 157
CAP 的作用机制 …… 158
7.2 *ara* 操纵子 …… 161
ara 操纵子的阻遏环 …… 161
ara 操纵子阻遏环的证据 …… 162
araC 的自主调节 …… 166
7.3 *trp* 操纵子 …… 166
色氨酸在 *trp* 操纵子负调控中的作用 …… 166
衰减作用对 *trp* 操纵子的调控 …… 167
衰减作用的失效 …… 167
7.4 核糖开关 …… 170
第 8 章 细菌转录机制的主要改变 …… 175
8.1 σ 因子的转换 …… 175
噬菌体感染 …… 175
孢子形成 …… 177
拥有多启动子的基因 …… 179
σ 因子的其他转换 …… 179
8.2 T7 噬菌体中所编码的 RNA 聚合酶 …… 180
8.3 λ 噬菌体感染 *E. coli* …… 181
λ 噬菌体的裂解繁殖 …… 182
溶源态的建立 …… 187
溶源期 *cI* 基因的自我调节 …… 189
λ 噬菌体感染后寄主命运的决定:裂解或溶源 …… 192
溶源菌的诱导 …… 194
第 9 章 细菌中 DNA-蛋白质的相互作用 …… 197
9.1 λ 阻遏物家族 …… 197
利用定点突变探测结合的专一性 …… 197
高分辨率分析 λ 阻遏物-操纵基因相互作用 …… 203
434 噬菌体阻遏物-操纵基因相互作用的高分辨率分析 …… 206
9.2 *trp* 阻遏物 …… 207
色氨酸的作用 …… 207
9.3 对蛋白质-DNA 相互作用的一般性

认识 …… 208
4 个不同碱基对的氢键形成能力 …… 208
多聚 DNA 结合蛋白的重要性 …… 209
9.4 DNA 结合蛋白:远程作用 …… 210
Gal 操纵子 …… 210
重复的 λ 操纵基因 …… 210
增强子 …… 212

第Ⅳ部分 真核生物的转录

第 10 章 真核生物的 RNA 聚合酶及其启动子 …… 216
10.1 真核生物 RNA 聚合酶的多种形式 …… 216
三类细胞核聚合酶的分离 …… 216
三类 RNA 聚合酶的作用 …… 217
RNA 聚合酶亚基结构 …… 219
10.2 启动子 …… 232
II 类启动子 …… 232
I 类启动子 …… 236
III 类启动子 …… 236
10.3 增强子与沉默子 …… 239
增强子 …… 239
沉默子 …… 241
第 11 章 真核生物中的通用转录因子 … 245
11.1 II 类转录因子 …… 245
II 类前起始复合物 …… 245
TFIID 的结构和功能 …… 247
TFIIB 的结构和功能 …… 257
TFIIH 的结构和功能 …… 261
中介物复合物和 RNA 聚合酶 II 全酶 …… 265
延伸因子 TFIIS …… 266
11.2 I 类转录因子 …… 269
核心结合因子 …… 270
UPE 结合因子 …… 271
SL1 的结构和功能 …… 271
11.3 III 类转录因子 …… 273
TFIIIA …… 273
TFIIIB 和 TFIIIC …… 274
TBP 的作用 …… 278
第 12 章 真核生物的转录激活因子 …… 282
12.1 激活因子的类型 …… 282
DNA 结合域 …… 282
转录激活域 …… 283
12.2 激活因子内 DNA 结合模体的结构 … 283
锌指结构 …… 283
GAL4 蛋白 …… 285
细胞核受体 …… 286
同源结构域 …… 287
亮氨酸拉链和螺旋-环-螺旋域 …… 288
12.3 激活因子各结构域的独立性 …… 290
12.4 激活因子的功能 …… 291
TFIID 的召集作用 …… 292
聚合酶全酶的召集 …… 292
12.5 激活因子间的相互作用 …… 296
二聚化作用 …… 296
远程作用 …… 297
复合增强子 …… 299
结构转录因子 …… 300
隔离子 …… 302
12.6 转录因子的调控 …… 305
辅激活因子 …… 305
激活因子的泛素化 …… 309
激活因子的 SUMO 修饰 …… 309
激活因子的乙酰化 …… 310
信号转导途径 …… 310
第 13 章 染色质结构及其对基因转录的影响 …… 315
13.1 组蛋白 …… 315
13.2 核小体 …… 316
30nm 纤丝 …… 319
染色质折叠的更高级结构 …… 321
13.3 染色质结构与基因活性 …… 323
组蛋白对 5S rRNA 基因转录的影响 …… 323
组蛋白对 II 类基因转录的影响 …… 324
核小体定位 …… 327
组蛋白乙酰化 …… 331
组蛋白去乙酰化 …… 332
染色质重建 …… 335
异染色质与沉默 …… 342
核小体与转录的延伸 …… 345

第Ⅴ部分 转录后加工

第 14 章 mRNA 加工 I:剪接 …… 350
14.1 断裂基因 …… 350

有关断裂基因的证据 …… 350
RNA 剪接 …… 352
剪接信号 …… 353
14.2 细胞核 mRNA 前体的剪接机制 … 354
分支中间体 …… 354
分支点信号 …… 356
剪接体 …… 357
剪接体的组装及功能 …… 369
定向、剪接位点的选择以及可变剪接 …… 373
剪接的调控 …… 383
14.3 RNA 的自我剪接 …… 385
I类内含子 …… 385
II类内含子 …… 390
第 15 章 mRNA 加工 II:加帽和多聚腺苷酸化 …… 394
15.1 加帽 …… 394
帽子的结构 …… 394
帽子的合成 …… 396
帽子的功能 …… 398
15.2 多聚腺苷酸化 …… 399
poly(A) …… 399
poly(A)的功能 …… 400
多聚腺苷酸化的基本机制 …… 403
多聚腺苷酸化信号 …… 404
前体 mRNA 的剪切和多聚腺苷酸化 …… 406
多聚腺苷酸酶 …… 413
poly(A)的更新 …… 414
15.3 mRNA 加工事件的协同运作 …… 416
帽子对剪接的作用 …… 416
poly(A)对剪接的作用 …… 417
Rpb1 的 CTD 与 mRNA 加工蛋白质的结合 …… 418
RNA 加工蛋白与 CTD 结合的变化与 CTD 磷酸化相关 …… 419
转录终止与 mRNA 3′端加工的偶联 …… 423
终止的机制 …… 424
多聚腺苷酸尾在 mRNA 转运中的作用 …… 427
第 16 章 其他 RNA 的加工过程 …… 431
16.1 核糖体 RNA(rRNA)的加工 …… 431
真核生物 rRNA 的加工 …… 431
细菌 rRNA 的加工 …… 434
16.2 转运 RNA(tRNA)的加工 …… 435
切开分离多顺反子前体 …… 435
成熟 5′端的形成 …… 435
成熟 3′端的形成 …… 436
16.3 反式剪接 …… 437
反式剪接机制 …… 437
16.4 RNA 编辑 …… 439
RNA 编辑的机制 …… 440
核苷酸脱氨基化编辑 …… 443
16.5 基因表达的转录后调控 …… 444
酪蛋白 mRNA 的稳定性 …… 444
转铁蛋白受体 mRNA 的稳定性 …… 445
RNA 干涉 …… 451
microRNA 和基因沉默 …… 463

第Ⅵ部分 翻 译

第 17 章 蛋白质翻译机制 I:起始 …… 471
17.1 细菌中翻译的起始 …… 471
tRNA 负载 …… 471
核糖体的解离 …… 472
30S 起始复合物的形成 …… 475
70S 起始复合物的形成 …… 481
细菌中的翻译起始总结 …… 483
17.2 真核生物翻译的起始 …… 484
起始的扫描模型 …… 484
真核生物的起始因子 …… 488
真核生物翻译起始概况 …… 488
17.3 翻译起始的调控 …… 496
细菌的翻译调控 …… 496
真核生物的翻译调控 …… 499
第 18 章 蛋白质翻译机制 II:延伸与终止 …… 512
18.1 多肽链合成及 mRNA 翻译的方向 …… 512
18.2 遗传密码子 …… 514
非重叠性密码子 …… 514
密码中无间隔 …… 514
三联密码 …… 515
破译密码 …… 516
密码子与反密码子间的异常碱基配对 …… 517
(几乎)通用的密码 …… 519
18.3 延伸机制 …… 520
延伸概述 …… 520
核糖体的 3 位点模型 …… 522

延伸步骤 1:氨酰-tRNA 与核糖体
A 位点的结合 …… 524
延伸步骤 2:肽键的形成 …… 529
延伸步骤 3:移位 …… 532
GTP 激酶和翻译 …… 534
EF-Tu 和 EF-G 的结构 …… 535
18.4 翻译的终止 …… 536
终止密码子 …… 536
终止密码子的抑制 …… 537
释放因子 …… 538
异常终止的处理 …… 541
利用终止密码子插入非寻常氨基酸 …… 545
18.5 翻译后 …… 546
新生蛋白质的折叠 …… 546
核糖体从 mRNA 的释放 …… 547
第 19 章 核糖体和转运 RNA …… 552
19.1 核糖体 …… 552
70S 核糖体的精细结构 …… 552
核糖体的组成 …… 555
核糖体的组装 …… 555
30S 亚基的精细结构 …… 557
50S 亚基的精细结构 …… 562
多聚核糖体 …… 566
19.2 转运 RNA …… 567
tRNA 的发现 …… 567
tRNA 的结构 …… 567
氨酰-tRNA 合成酶对 tRNA 的识别:第二遗传密码 …… 570
氨酰-tRNA 合成酶的校正和编辑 …… 575

第Ⅶ部分 DNA 复制、重组和转座

第 20 章 DNA 复制 I:基本机制与酶学 …… 579
20.1 DNA 复制的一般特征 …… 579
半保留复制 …… 579
半不连续复制 …… 580
DNA 合成的引发 …… 583
双向复制 …… 585
单向复制 …… 588
滚环复制 …… 588
20.2 DNA 复制的酶学 …… 589
*E. coli*中的 3 种 DNA 聚合酶 …… 589
复制保真性 …… 593
真核生物的多种 DNA 聚合酶 …… 594
链的分离 …… 595
单链 DNA 结合蛋白 …… 597
拓扑异构酶 …… 598
20.3 DNA 损伤与修复 …… 601
碱基的烷基化修饰引起的 DNA 损伤 …… 602
紫外线辐射引起的 DNA 损伤 …… 603
γ 射线及 X 射线引起的 DNA 损伤 …… 603
DNA 损伤的直接修复 …… 604
切除修复 …… 605
真核生物的双链断裂修复 …… 611
错配修复 …… 613
人类细胞错配修复系统的失效 …… 613
未修复 DNA 损伤的处理 …… 614
第 21 章 DNA 复制 II:详细机制 …… 622
21.1 复制的起始 …… 622
E. coli DNA 复制的引发 …… 622
真核生物 DNA 复制的引发 …… 624
21.2 延伸 …… 628
复制的速度 …… 628
pol III 全酶与复制的进行性 …… 629
21.3 复制的终止 …… 642
解连接:解开缠绕的子代 DNA 分子 …… 642
真核生物 DNA 复制的终止 …… 644
第 22 章 同源重组 …… 656
22.1 同源重组的 RecBCD 途径 …… 657
22.2 RecBCD 途径的实验证据 …… 659
RecA …… 659
RecBCD …… 662
RuvA 和 RuvB …… 664
RuvC …… 667
22.3 减数分裂重组 …… 669
减数分裂重组的机制:综述 …… 669
双链 DNA 断裂 …… 669
DSB 处单链末端的生成 …… 675
22.4 基因转换 …… 676
第 23 章 转座 …… 679
23.1 细菌转座子 …… 679
细菌转座子的发现 …… 679
插入序列:最简单的细菌转座子 …… 679
更复杂的转座子 …… 680
转座的机制 …… 682

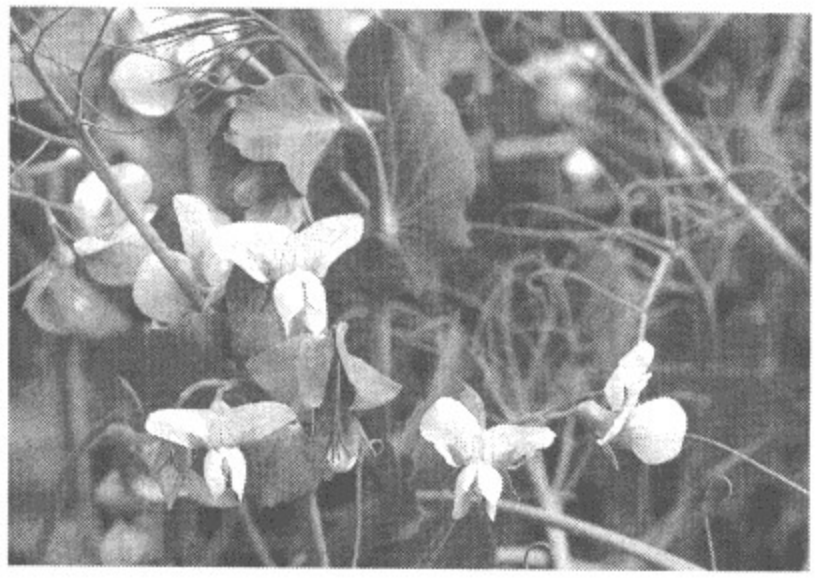
23.2 真核生物的转座子 …… 684
第一例转座因子:玉米的 *Ds*/*Ac* 系统 …… 684
P 因子 …… 686
23.3 免疫球蛋白基因的重排 …… 686
重组信号 …… 688
重组酶 …… 689
V(D)J 重组机制 …… 689
23.4 反转录转座子 …… 691
反转录病毒 …… 691
反转录转座子 …… 696

第Ⅷ部分 基因组

第 24 章 基因组学、蛋白质组学和生物信息学 …… 704
24.1 图位克隆:基因组学介绍 …… 704
图位克隆的传统手段 …… 704
鉴定与人类疾病相关的突变基因 …… 706
24.2 基因组测序 …… 711
人类基因组计划 …… 714
用于大规模基因组计划的载体 …… 714
逐步克隆策略 …… 716
鸟枪法测序 …… 718
测序的标准 …… 719
人类基因组测序 …… 719
其他脊椎动物基因组 …… 725
最小的基因组 …… 727
生命条形码 …… 727
24.3 基因组学的应用 …… 728
转录组学 …… 728
基因组功能谱 …… 736
单核苷酸多态性:药物基因组学 …… 742
24.4 蛋白质组学 …… 743
蛋白质分离 …… 744
蛋白质分析 …… 744
蛋白质的相互作用 …… 745
24.5 生物信息学 …… 748
从哺乳动物基因组中发现调控基序 …… 748
如何使用数据库 …… 750

分子生物学专业词汇表 …… 756
参考文献 …… 792
索引 …… 810

第1章　分子生物学简史



豌豆植株，孟德尔的遗传学研究实验材料。 *Holt Studios International (Nigel Cattlin) /Photo Researchers, Inc.*

什么是分子生物学？这个术语有多种定义。广义的定义是在分子水平解释生物学现象，但这种定义难以与生物化学区分。另一种更为严格，因而也更为有用的定义是指在分子水平研究基因的结构和功能，本书将从分子水平解释基因及其活性（activity）。

分子生物学源于遗传学和生物化学。本章从19世纪中期孟德尔最早的遗传学实验开始，回顾这门交叉学科的早期发展史。第2、3章将充实这一简要的发展轮廓。由于早期的遗传学家还不知道基因的分子本质，根据定义对基因的早期研究不能归入分子生物学或分子遗传学，代之于称其为传递遗传学（transmission genetics），研究遗传性状从亲本向子代的传递。到1944年基因的化学组成搞清楚后，将基因作为分子进行研究才成为可能，分子生物学得以诞生。

1.1　传递遗传学

1865年，孟德尔（Gregor Mendel）（图1.1）发表了对豌豆7个遗传性状的研究结果。在孟德尔的研究之前，科学家认为遗传是以双亲性状在子代混合的方式发生的，但孟德尔推断遗传是颗粒式的（particulate），即每个亲本将颗粒或者遗传单位传给子代。我们现在称这些颗粒为**基因**（gene）。通过仔细计算具有一定**表型**（phenotype）或可观察特征（如黄色种子、白色花朵等）的后代植株数，孟德尔

图1.1　孟德尔。（*Source*: Courtesy of Dept. of Library Services/American Museum of Natural History, Neg. no. 219467.）

得出一些重要结论。表型（phenotype）一词来源于相同的希腊语词根 *phenomenon*，意思是外表。高秆豌豆表现高秆表型。表型一词也指生物体的所有可观察的特征。

孟德尔的遗传定律

孟德尔观察到一个基因能以不同的形式存在，叫做**等位基因**（allele）。例如，豌豆种子有黄色或绿色，控制种子颜色的一个等位基因产生黄色种子，另一个则产生绿色种子。相对于**隐性**（recessive）等位基因，另一个等位基因是**显性的**（dominant）。孟德尔证明黄色种子的等位基因是显性的，用绿豌

豆与黄豌豆杂交时，子一代（first filial generation）（F_1）种子全部为黄色，但是让这些子一代（F_1）黄豌豆自交时，又出现一些绿豌豆。在子二代（F_2）中黄色与绿色的比例非常接近 3∶1。

filial 一词来源于拉丁文 *filius*（儿子）和 *filia*（女儿），因此，F_1 代包括了亲本所有后代的信息。子二代（F_2）是子一代（F_1）个体的后代。

孟德尔相信绿色等位基因一定被保存在 F_1 代，虽然它不影响 F_1 种子的颜色。其解释是：至少对他正在研究的性状来讲，每个亲本植株携带两拷贝的基因，即亲本是**二倍体**（diploid）。根据这个概念，**纯合子**（homozygote）具有两拷贝的相同等位基因，两个黄色等位基因或两个绿色等位基因。**杂合子**（heterozygote）的每个等位基因只有一拷贝。父本与母本在第一次杂交之前都是纯合子，所产生的 F_1 代豌豆都是杂合子。此外，孟德尔推断生殖细胞只含一个拷贝的基因，即生殖细胞是**单倍体**（haploid）。因此，纯合子产生只含一个等位基因的生殖细胞或**配子**（gamete），而杂合子可产生含任一等位基因的配子。

黄豌豆与绿豌豆杂交时，黄色亲本贡献带有黄色基因的配子，绿色亲本贡献带有绿色基因的配子。所有 F_1 代豌豆都有一个黄色等位基因和一个绿色等位基因，它们根本没有丢失绿色基因，但因为黄色是显性基因，因而所有种子都显黄色。当这些杂合豌豆自交时，产生相同数目的黄色配子和绿色配子，使绿色表型得以重现。

现在来分析一下上述过程是怎样发生的。假设有两个袋子，每只袋子里有相同数目的黄色和绿色弹子。如果每次从一个袋子里取出一颗弹子，同时从另一个袋子里也取出一颗弹子配成一对。可得出以下结果：1/4 黄/黄、1/4 绿/绿、其余 1/2 黄/绿。豌豆的黄色和绿色等位基因以相同的方式进行配对。黄色是显性基因，只能看到 1/4 的子代（绿/绿）是绿色，其余 3/4 为黄色，因为它们至少有一个黄色等位基因。因此，在子二代（F_2）黄豌豆与绿豌豆的比例是 3∶1。

孟德尔通过对 7 个不同性状的研究还发现这些基因是相互独立发挥作用的。因此，两对等位基因组合（如黄色或绿色豌豆与饱满或皱缩豌豆。其中，黄色和饱满为显性，绿色和皱缩为隐性）所产生黄色/饱满、黄色/皱缩、绿色/饱满、绿色/皱缩的比率分别为 9∶3∶3∶1。服从孟德尔所发现的简单规律的遗传学称为**孟德尔遗传**（Mendelian inheritance）。

小结 基因以不同形式或等位基因存在。一对等位基因中，一个是显性的，另一个是隐性的。因此，杂合子含有基因的两个不同等位基因，通常表现由显性等位基因决定的性状，隐性等位基因并没有丢失，当与另一个隐性等位基因配对成纯合子时可发挥作用。

遗传的染色体理论

孟德尔研究的价值在当时未受到其他科学家的理解和重视，直到 1900 年三位植物学家各自独立研究都得出与孟德尔相似的结论后，人们才重新发现孟德尔的研究价值。

1900 年之后，多数遗传学家都接受了基因的颗粒特性，遗传学开始发展起来。对染色体性质的不断了解促使遗传学家更容易接受孟德尔理论，染色体理论始于 19 世纪后半叶。

孟德尔推测配子只携带每对等位基因中的一个。如果染色体携带基因，那么在配子体中染色体的数目应该减半，事实确实如此。因此，染色体是承载基因的不连续物理实体。

染色体携带基因的观点就是遗传的**染色体理论**（chromosome theory of inheritance），这是遗传学思维的重要新阶段。基因不再是无实体的因子，而是细胞核中的可见物质。但是一些遗传学家，特别是摩尔根（Thomas Hunt Morgan）（图 1.2），对这一观点持怀疑态度。有意思的是 1910 年摩尔根本人却为染色体学说提供了第一个决定性的证据。

图 1.2 摩尔根。（*Source*：National Library of Medicine.）

摩尔根以果蝇（*Drosophila melanogaster*）为实验材料。果蝇具有个体小、世代周期短、后代数量多等优点，在许多方面比豌豆更适合做遗传学研究。将红眼果蝇（显性）与白眼果蝇（隐性）杂交，大部分 F_1 代为红眼的。用 F_1 代的雄性红眼与其红眼姐妹杂交后，产生约 1/4 的雄性白眼，而不是雌性

白眼。换句话说，眼色表型是**性连锁的**（sex-linked），随性别一起遗传。为什么会这样呢？

我们现在知道，性别和眼色一起传递是因为控制这些性状的基因都位于同一条 X 染色体上。大部分染色体叫做**常染色体**（autosome），在个体中成对出现，但 X 染色体是一种**性染色体**（sex chromosome）。雌果蝇有两拷贝 X 染色体，而雄果蝇只有一拷贝。摩尔根不愿意得出这一结论，直到同年在另外两个表型（残翅和黄体色）中也观察到相同的性连锁现象，才足以说服他接受遗传染色体理论的正确性。

在结束这个话题之前，还要指出两点。第一，每个基因在染色体上都有其位置或**基因座**（locus）。图 1.3 是假想的染色体及其 3 个基因 A、B、C 的位置。第二，像人类这样的二倍体生物，所有染色体（性染色体除外）通常是两拷贝的，意味着它们的大部分基因是两拷贝的。如果这两个拷贝是相同的等位基因，则生物体是**纯合的**（homozygous），否则是**杂合的**（heterozygous）。图 1.3b 显示二倍体染色体在（*Aa*）基因座有不同的等位基因，在另外两个基因座（*BB* 和 *cc*）有相同的等位基因，其**基因型**（genotype）或等位基因构成是 *AaBBcc*。该生物的两条染色体在座位 *A* 有两个不同的等位基因（*A* 和 *a*），所以在 *A* 座位是杂合的（希腊文：*hetero*，意思是不同），在 *B* 座位有相同的显性等位基因 *B*，所以在 *B* 座位是显性纯合的（希腊文：*homo*，意思是相同）。两条染色体的 *C* 座位有相同的隐性等位基因 *c*，因此在 *C* 座位是隐性纯合的。由于 *A* 相对于 *a* 是显性的，所以该个体在 *A*、*B* 座位是显性表型，在 *C* 座位是隐性表型。

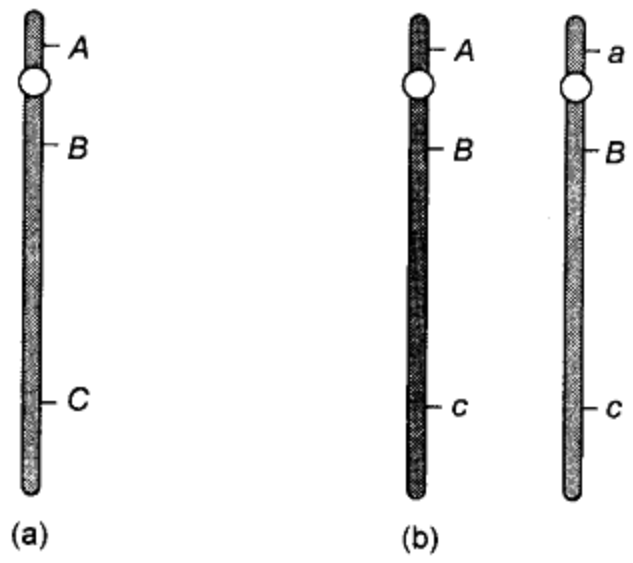


图 1.3　基因在染色体上的位置。（a）染色体示意图，给出了 3 个基因的座位 *A*、*B* 和 *C*；（b）双倍体成对染色体示意图，显示了 3 个基因 *A*、*B*、*C* 各自的座位，以及在每个座位的基因型（*A* 或 *a*；*B* 或 *b*；*C* 或 *c*）。

讨论果蝇的表型变异使我们有机会介绍另一个重要遗传学概念：**野生型**（wild-type）和**突变型**（mutant）。野生型表型是最普通（至少是普遍接受）的标准表型。为避免野生个体自然就是野生型的错觉，有些遗传学家喜欢用**标准型**（standard type）。在果蝇中，红眼和全翅是野生型。白眼（*white*，*w*）和残翅（*miniature*，*m*）基因突变分别导致了白眼和残翅果蝇。突变的等位基因通常是隐性的，但不能一概而论。

遗传重组和遗传图谱

在遗传实验中位于不同染色体上的基因独立遗传，而同一染色体上的基因如残翅（*m*）和白眼（*w*）表现出连锁。但是，在同一染色体上的基因通常并不显示完全的**遗传连锁**（genetic linkage）。这一现象是摩尔根在检测性连锁基因的行为时发现的。例如，*m* 和 *w* 都在 X 染色体上，但是它们的后代中只有 65.5%保留连锁，其他后代则产生了在亲本中不存在的新等位基因组合，称为**重组体**（recombinant）。

重组体是如何产生的呢？答案在 1910 年就有了。对减数分裂（meiosis）（配子形成）的显微观察发现**同源染色体**（homologous chromosome）（携带相同基因，或相同基因的等位基因）之间发生了**交换**（crossing），导致两个同源染色体之间的基因交换。上例中，卵子形成过程中带有 *m* 和 *w* 等位基因的 X 染色体与另一个带有红眼正常翅等位基因的 X 染色体发生交换（图 1.4）。由于产生了等位基因的新组合，所以我们称这个过程为重组（recombination）。

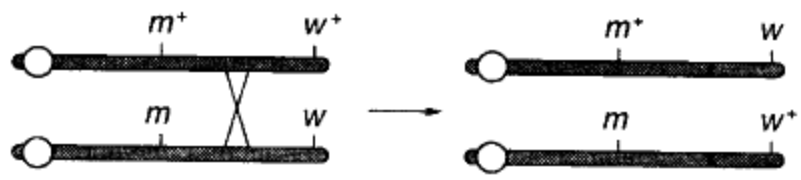


图 1.4　果蝇中的重组。显示雌果蝇的两条 X 染色体。一条携带两个野生型基因，产生正常翅（m^+）和红眼（w^+）。另一条携带两个突变基因，产生残翅（*m*）和白眼（*w*）。卵形成过程中，如交叉直线所示两条染色体的两个基因之间发生重组或联会。结果产生两条重组染色体，一个是 m^+w，另一个是 mw^+。

摩尔根推测基因在染色体上呈线性排列，就像用线穿起来的珠子。根据他对重组的了解，摩尔根认为，同一染色体上的两个基因相隔越远，它们被重组的可能性越大。因为相隔远的基因之间有更大的空间产生交换。A. H. Sturtevant 拓展了这个假

设，认为一条染色体上两个基因相隔的距离与它们之间的重组率有一定的数学关系。Sturtevant 收集了果蝇重组的数据来支持其假设，并建立了至今仍在使用的**遗传作图**（genetic mapping）技术的基本原理。简单地讲，如果两个位点重组的频率是 1%，则定义它们相距一个**厘摩**（centimorgan）（以摩尔根命名）。到 20 世纪 30 年代，该原理也被发现适用于其他**真核生物**（eukaryote）（具有细胞核的生物），包括链孢霉、豌豆、玉米甚至人，以及无细胞核的**原核生物**（prokaryote，遗传物质不限制在核中）。

重组的物理学证据

Barbara McClintok（图 1.5）和 Harriet Creighton 在 1931 年提供了重组的直接物理学证明。通过对玉米染色体的显微观察，他们检测到特定染色体上两个易鉴别特征（一端有节，另一端是长的伸展）间的重组。而且这种物理重组也能从遗传学上检测到。由此建立了染色体的一段区域与某个基因相对应的直接关系。不久，Curt Stern 在果蝇上也观察到同样现象。因此，在动植物中都能对重组做出物理学和遗传学上的检测。

图 1.5　Barbara。（*Source*：Bettmann Archive/Corbis.）

小结　遗传的染色体理论指出基因在染色体上呈线性分布。有些性状可共同遗传的原因是控制这些性状的基因位于同一条染色体上。然而，在减数分裂期两条同源染色体之间的重组使亲代的等位基因混杂产生了非亲代重组。同一染色体上的两个基因相距越远，产生重组的可能性越大。

1.2　分子遗传学

通过以上讨论我们了解到关于基因传递、染色体上基因定位的重要内容，但这些工作没有告诉我们基因由什么构成、怎样发挥作用。这些问题属于分子遗传学领域，该领域也根植于孟德尔时代。

DNA 的发现

1869 年，Friedrich Miescher（图 1.6）在细胞核中发现了一种混合物，并将其称为核素（nuclein），其主要成分是**脱氧核糖核酸**（deoxyribonucleic acid）（DNA）。到 19 世纪末，化学家已经知道了 DNA 及相关化合物**核糖核酸**（ribonucleic acid）（RNA）的一般结构。两者都是长的聚合物，即由小分子化合物核苷酸组成的链状结构。每个核苷酸由一个糖、一个磷酸基团和一个碱基组成。糖基彼此通过磷酸二酯键相互连接形成核苷酸链。

图 1.6　Friedrich。（*Source*：National Library of Medicine.）

基因的组成　随着遗传染色体理论被广泛接受，遗传学家一致认为染色体必定由某种多聚物组成，以符合其作为一连串基因的角色，但这种多聚物是什么呢？有 3 种选择：DNA、RNA 和蛋白质。蛋白质是 Miescher 所发现核素中的另一种主要成分，是由**氨基酸**（amino acid）通过**肽键**（peptide bond）连接而成的链，单个蛋白质链称为**多肽**（polypeptide）。

1944 年，Oswald Avery（图 1.7）及同事在 Frederick Griffith 的细菌遗传性状转移实验的基础上证明 DNA 就是携带遗传物质的多聚物（见第 2 章）。他们研究的遗传性状是**毒性**（virulence），引起致死性感染的能力。将死亡的毒性细胞与活的无毒性（非致死性的）细胞混合就能够进行毒性转化。使受体细胞由无毒性转变成有毒性的物质很可能是毒性基因，因为受体细胞可把这种性状传递给后代。

有待弄明白的是死的毒性细胞中转化源的化学本质。Avery 等用几种化学和生物化学方法对转化源进行检测，结果表明其性质是 DNA，而不是 RNA 或者蛋白质。

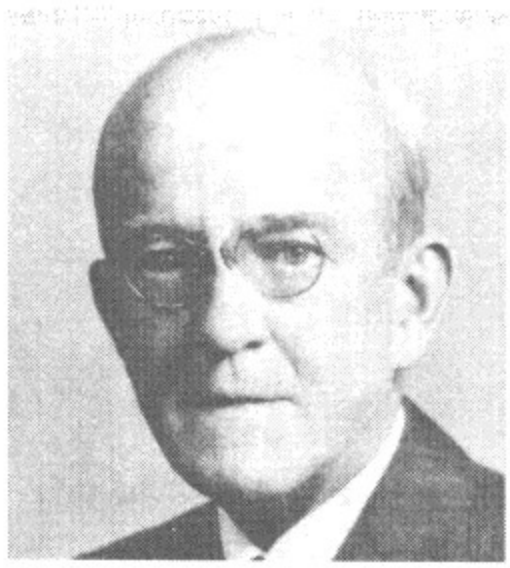


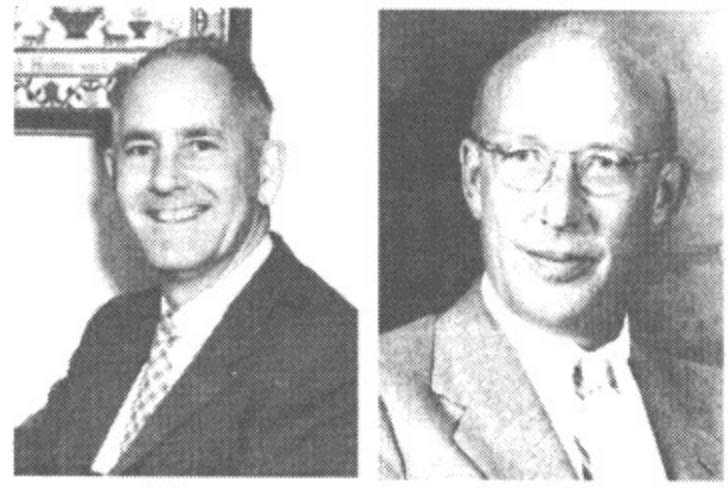
图 1.7　Oswald。（*Source*：National Academy of Sciences.）

基因和蛋白质之间的关系

分子遗传学的另一个重要问题是基因如何发挥作用？为了回答这个问题，需追溯到 1902 年。这一年 Archibald Garrod 注意到人类尿黑酸症（alcaptonuria）似乎是一种孟德尔隐性性状，可能由一个缺陷或突变的基因引起。其主要症状是患者尿液里有黑色素沉积，Garrod 认为这源于某个生化途径中某种中间产物的不正常积累。

此时，生物化学家已经证实所有生物体内都发生着不计其数的、由称之为**酶**（enzyme）的蛋白质所加速或催化的化学反应。许多化学反应是按顺序进行的，即一个产物是下一个反应的起始物或底物。这种反应过程被称为**途径**（pathway），反应途径内的产物或底物被称为**中间产物**（intermediate）。Garrod 认为尿黑酸症患者体内某种中间产物积累到了异常高的水平，因为将其正常地转化为下一产物的酶缺失了。结合尿黑酸症是孟德尔隐性性状，Garrod 提出一个缺陷的基因导致一个缺陷的酶。换句话说：一个基因产生一种酶。

Garrod 的结论部分是推测的，他其实不知道哪个缺陷的酶与尿黑酸症有关。这个问题留给了 George Beadle 和 E. L. Tatum（图 1.8）来证明基因与酶之间的关系。他们以链孢霉（*Neurospora*）为实验材料，其优点是不受自然突变限制，且可以使用**突变剂**（mutagen）引入突变，然后观察这些突变在生化途径中的效果。他们发现可以在很多时刻引入突变获得链孢霉突变体，然后把缺陷追踪到生化途径的某个步骤上，再具体到某个酶（见第 3 章）。通过加入该缺失酶在正常情况下产生的中间产物使其恢复正常生长可实现这一策略。通过疏通阻塞的地方，可以发现酶的位置。他们的遗传学实验显示，单个基因参与了反应。因此，缺陷基因造成了缺陷（或缺失）酶。换句话说，一个基因产生一种酶。这就是一个基因/一个酶假说。但是这个假说由于至少 3 方面原因是不完全正确的。①一个酶可以由多条多肽链组成，而一个基因携带的信息只能形成一条多肽链。②多数基因携带的信息所产生的多肽并不都是酶。③可以看到，有些基因的产物不是多肽而是 RNA。对这个假说的现代叙述应该是：大多数基因含有产生一条多肽的信息。



(a)　　(b)

图 1.8　(a)George Beadle；(b) E. L. Tatum.（*Source*：(a，b) AP/Wide World Photos.）

基因的活性

现在让我们回到手头的问题：基因如何发挥作用？这其实不是一个单一问题，因为基因不止做一件事情。首先，它们被忠实地复制；其次，它们指导 RNA 和蛋白质的产生；第三，它们积累突变，允许进化。让我们简要地了解一下基因的这 3 种活性。

基因怎样被复制　DNA 如何被忠实复制？回答这个问题前，我们需要知道 DNA 分子在染色体上的总体结构。1953 年，沃森（James Watson）和克里克（Francis Crick）（图 1.9）通过其他实验室的

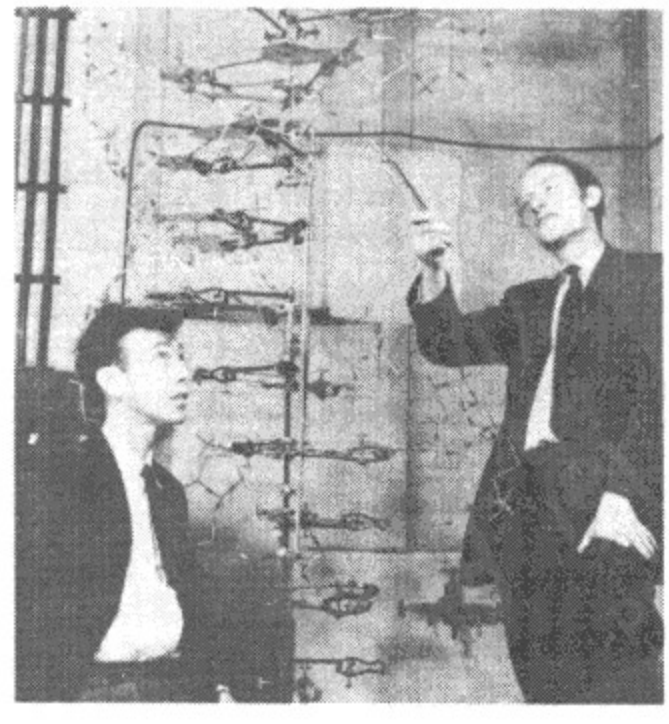


图 1.9　沃森（左）和克里克。（*Source*：Corbis/Bettmann Archive.）

化学和物理实验数据，主要是 Rosalind Franklin 和 Maurice Wilkins（图 1.10）的 X 射线衍射数据，建立模型并提供了答案。

(a) (b)

图 1.10 (a) Rosalind Franklin; (b) Maurice Wilkins。（*Sources*：(a) From *The Double Helix* by James D. Watson, 1968, Atheneum Press, NY. © Cold Spring Harbor Laboratory Archives. (b) Courtesy Professor M. H. F. Wilkins, Biophysics Dept., King's College, London.）

沃森和克里克推测 DNA 是一种**双链螺旋**（double helix）——两条 DNA 链相互缠绕。更重要的是，每条链的碱基都位于螺旋的内部，一条链上的碱基以一种特殊的方式与另一条链上的碱基配对。DNA 中只有 4 种不同的碱基：腺嘌呤、鸟嘌呤、胞嘧啶和胸腺嘧啶，分别缩写成 A、G、C、T。无论在一条链的任何地方发现一个 A，在另一条链上就总会有一个 T；无论在一条链的任何地方发现一个 G，在另一条链上就总会有一个 C，就是说这两条链是互补的。如果我们知道一条链的碱基顺序，就自然会知道另一条链的碱基顺序。这种互补性使 DNA 能够被忠实地复制。两条链分离时，酶利用旧链作为模板再根据沃森-克里克碱基配对原则（A 与 T，G 与 C）形成新的互补链，这称为**半保留复制**（semiconservative replication），因为亲代双螺旋中的一条链被保留在每个子代双螺旋中。1958 年 Matthew Meselson 和 Franklin Stahl（图 1.11）证明在细菌中 DNA 复制遵守半保留方式（见第 20 章）。

基因如何指导多肽的产生 基因表达是基因产物（RNA 或者多肽）生产的过程。分别叫做**转录**（transcription）和**翻译**（translation）的两个步骤是 DNA 基因指导产生多肽所必需的。在转录步骤中，叫做 RNA 聚合酶的一种酶生产出其中一条 DNA 链的副本。这个副本不是 DNA，而是它的近亲 RNA。在翻译过程中，这个 RNA（messenger RNA，信使 RNA，mRNA）携带遗传指令到达细胞内叫做**核糖体**（ribosome）的蛋白质工厂中。核糖体“阅读”

(a) (b)

图 1.11 (a) Matthew Meselson; (b) Franklin Stahl。（*Sources*：(a) Courtesy Dr. Matthew Meselson. (b) Cold Spring Harbor Laboratory Archives.）

mRNA 上的**遗传密码**（genetic code）并根据它的指令生产出蛋白质。

实际上，核糖体已经含有 RNA 分子，称为**核糖体 RNA**（ribosomal RNA，rRNA）。克里克最初认为是位于核糖体上的这个 RNA 携带了来自基因的信息。根据这个理论，每个核糖体只能生产一种由其 rRNA 编码的蛋白质。François Jacob 和 Sydney Brenner（图 1.12）则认为核糖体不是特异性翻译机器，可根据访问核糖体的 mRNA 的指令，产生无数不同的蛋白质，实验已证明这个观点是正确的（见第 3 章）。

(a) (b)

图 1.12 (a) François Jacob; (b) Sydney Brenner。（*Source*：(a, b) Cold Spring Harbor Laboratory Archives.）

遗传密码的本质是什么？Marshall Nirenberg 和 Har Gobind Khorana（图 1.13）在 20 世纪 60 年代的早期通过各自独立的不同方法破译了遗传密码（见第 18 章）。他们发现 3 个碱基组成一个密码，称之为**密码子**（codon），代表一个氨基酸。在 64 个可能的三联密码子中，有 61 个编码氨基酸，其他 3 个是终止信号。

核糖体每次扫描一个 mRNA 的 3 个碱基，并将相应的氨基酸送到增长的蛋白质链上连接起来。当

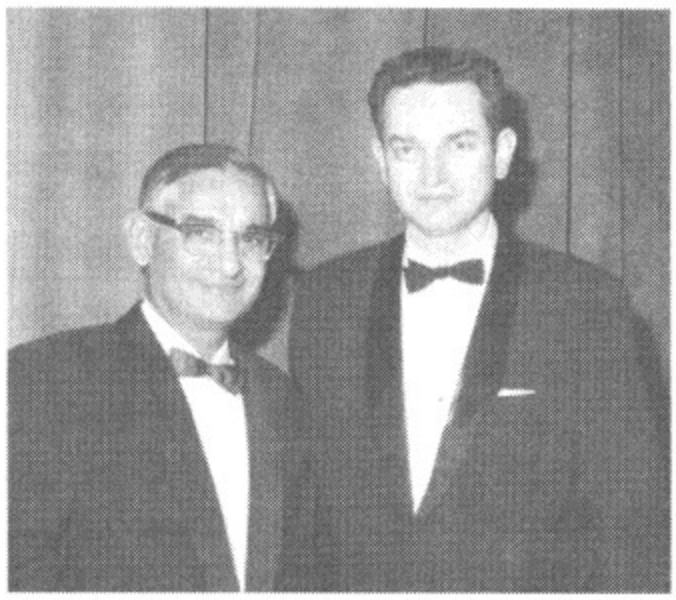


图 1.13　Har Gobind Khorana（左）和 Marshall Nirenberg（右）。（*Source*：Corbis/Bettmann Archive.）

它们到达终止密码子时，就释放出完整的蛋白质。

基因如何积聚突变？　基因以多种方式发生改变。最简单的是由一个碱基变为另一个碱基。例如，密码子 GAG（对应谷氨酸）变为 GTG 就成了缬氨酸密码子。由这个突变基因产生的蛋白质在本应是谷氨酸的位置变成了缬氨酸。这可能是数百个氨基酸中的一个变化，但是它能产生深刻的影响。人血红蛋白基因的一个碱基改变，可导致镰状细胞贫血症的遗传紊乱。

基因还会遭受更深刻的改变，如大片段 DNA 的删除或插入。DNA 片段甚至能从一个位点转移到另一个位点。基因所经受的变化越剧烈，就越有可能完全失活。

基因克隆　近几年来，遗传学家学会了分离基因，并把它们放入新的生物体内，通过一套叫做**基因克隆**（gene cloning）的技术扩增基因。克隆的基因不仅给分子生物学家提供了大量的研究材料，还可以被诱导产生相应的蛋白质产物。有些产物（如人胰岛素或凝血因子）都是非常有用的。克隆的基因还能被移植到植物、动物（包括人类）中。这些移植的基因能改变生物受体的性状，因此，它们可以为农业和治疗人类遗传疾病提供强有力的工具。我们将在第 4 章详细研究基因克隆。

小结　所有基因都是由排列在双螺旋上的 DNA 构成。这种结构解释了基因是怎样参与 3 种主要活动的：复制、携带信息、积累突变。基因中两条 DNA 链的互补本质可以使它们通过分离并作为模板组装两条新的互补链而被忠实地复制。基因中的核苷酸序列就是携带合成 RNA 信息的遗传密码。把信息携带到蛋白质合成场所核糖体的 RNA 大部分是 mRNA。最终的结果是根据基因指令生产一个新的蛋白质链。碱基序列的一个改变就构成突变，突变可能导致基因的蛋白质产物中氨基酸序列改变。基因可以被克隆，使分子生物学家能够收获大量的克隆基因的产物。

1.3　生命的 3 个领域

20 世纪早期，科学家把所有生物分为两界：动物界和植物界。细菌被认为是植物，这就是为什么我们仍然把小肠内的细菌称作肠菌群（flora）。但是 20 世纪中期以后，这种分类系统被放弃而采用五界分类系统（five-kindom system），除了植物和动物之外还包括细菌、真菌和原生生物。

20 世纪 70 年代后期，Carl Woese（图 1.14）对许多不同生物的核糖体 RNA 基因进行测序研究并得出一个惊人的结论：一类被划分为细菌的生物，其 rRNA 基因与真核生物的相似性高于典型的细菌，如大肠杆菌。因此，Woese 将这些生物命名为古菌（*archaebacteria*），以区别于真正的细菌或真细菌（*eubacteria*）。然而，随着越来越多的分子证据的积累，可以清楚地看出除了表面的相似之外，古菌并不是真正的细菌，它们代表了一个截然不同的生命领域。所以 Woese 将它们的名字改为古菌。现在我们按 3 个领域认识生命：**细菌**（bacteria）、**真核生物**（eukaryota）和**古菌**（archaea）。与细菌一样，古菌是**原核生物**（prokaryote）——无细胞核生物，但是它们的分子生物学实际上与真核生物更接近。

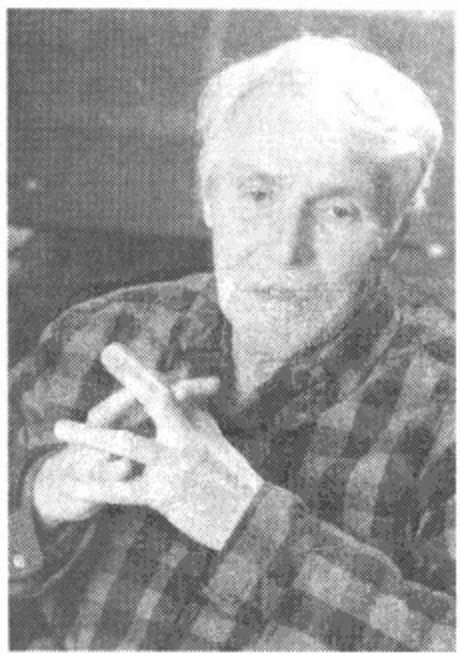


图 1.14　Carl Woese。（*Source*：Courtesy U. of Ill at Urbana Champaign.）

古菌生活在地球最恶劣的地区。有些是**嗜热生物**（thermaophile）（“好热生物”），生活在看上去似乎不能忍受的、温度超过 100℃ 的深海地热出口附

近，或生活在像黄石国家公园的高温温泉中。有些是**嗜盐生物**（halophile）（“好盐生物”），能够忍受使其他生命脱水或致死的高盐浓度。还有些是**甲烷细菌**（methanogen）（“甲烷生产者”），生活在牛胃中，这就是为什么牛胃是甲烷的一个重要来源的原因。

本书将主要涉及前两个生命领域，因为它们被研究得最多。但是在第 11 章我们将学习古菌转录的有趣细节。在第 24 章我们将了解到 *Methanococcus jannaschii*，一种古菌，是第一批进行基因组测序的生物之一。

小结 所有的生物被划分为 3 个领域：细菌、真核生物和古菌。尽管表面上古菌与细菌类似，但其分子生物学特性更相似于真核生物。

以上总结归纳了分子生物学发展的主要大事记。表 1.1 回顾了其中的一些里程碑。尽管分子生物学是一门非常年轻的学科，但却有异常丰富的历史。分子生物学家现在正以爆炸性的速度为这门学科增加新的知识。分子生物学的发展速度和它的技术威力使许多评论家称其为一场革命。由于在接下来的几十年中，医学和农业上的一些最重要变化可能依赖于分子生物学家的基因操作，所以这场革命将以不同方式触及每个人的生活。因此，你正要开始学习的这门课程不仅令人神往、精致优美而且具有实践重要性。哈佛大学名誉退休化学教授 F. H. Westheimer 说得好：“近 40 年最伟大的知识革命可能已在生物学发生。今天，一个不懂得点分子生物学知识的人能被认为是受过教育的人吗?”很高兴，学完这门课程后你们将懂得远不止一点。

表 1.1 分子生物学年代记事

1859 年	Charles Darwin	出版《物种起源》
1865 年	Gregor Mendel	分离和独立分配定理
1869 年	Friedrich Miescher	发现 DNA
1900 年	Hugo de Vries，Carl Correns，Erich von Tschermak	孟德尔定理再次被发现
1902 年	Archibald Garrod	首次提出一种人类疾病的遗传学起因
1902 年	Walter Sutton，Theodor Bovein	提出染色体理论
1910 年，1916 年	Thomas Hunt Morgan，Calvin Bridges	证明基因在染色体上
1913 年	A. H. Sturtevant	构建遗传图谱
1927 年	H. J. Muller	用 X 射线引入突变
1931 年	Harriet Creighton，Barbara McClintock	获得重组的物理证据
1941 年	George Beedle，E. L. Tatum	提出一个基因/一个酶假说
1944 年	Oswald Avery，Colin Mcleod	证明基因由 DNA 组成
1953 年	James Watson，Francis Crick，Rosalind Frankin，Maurice Willkins	确定 DNA 结构
1958 年	Matthew Meselson，Franklin Stahl	证明 DNA 的半保留复制
1961 年	Sydney Brenner，François Jacob，Matthew Meselson	发现信使 RNA
1966 年	Marshall Nirenberg，Gobind Khorana	完成遗传密码破解
1970 年	Hamilton Smith	发现限制性内切核酸酶在专一位点切割 DNA，使切补 DNA 变得容易，促进了 DNA 克隆技术
1972 年	Paul Berg	获得第一个体外重组 DNA
1973 年	Herb Boyer，Stanly Cohen	第一次用质粒克隆 DNA
1977 年	Walter Gilbert 和 Frederick Sanger	找出测定 DNA 碱基序列的方法
1977 年	Frederick Sanger	测定出 ΦX174 病毒的完整基因组序列
1977 年	Phillip Sharp，Richard Roberts 等人	发现基因中有间断（内含子）
1990 年	James Watson 及多人	公布人类基因组计划，绘制基因组图谱最终测定碱基序列
1993 年	Huntington 病研究组	鉴定出 Huntington 症的致病基因
1995 年	Craig Venter，Hamilton Smith	测定了 *Haemophilus influenzae* 和 *Mycoplasma genitalium* 两个细菌的基因组碱基序列，对自由生活有机体基因组的第一次测定

续表

1996 年	众多研究者	测定酿酒酵母 *Saccharomyces cerevisiae* 基因组，对真核生物基因组的第一次测定
1997 年	Ian Wilmut 及同事	从成体绵羊体细胞获得克隆羊（Dolly）
1998 年	Andrew Fire 及同事	发现 RNAi 通过降解与侵入的双链 RNA 具有相同序列的 mRNA 而起作用
2003 年	众多研究者	报道完成人类基因组序列测序
2005 年	众多研究者	报道我们的最近亲猩猩基因组草图

总结

基因以等位基因的几种不同形式存在。在杂合子中，隐性等位基因会被显性等位基因遮盖但并没有丢失，在携带有 2 个隐性等位基因的纯合子中会被再次表达。

基因在染色体上呈线性分布。因此，位于同一条染色体上的基因所控制的性状可被同时遗传。然而，在减数分裂期两条同源染色体之间会发生重组，由此产生携带非亲代重组等位基因的配子。在同一条染色体上的两个基因相距越远，产生这种重组的可能性就越大。

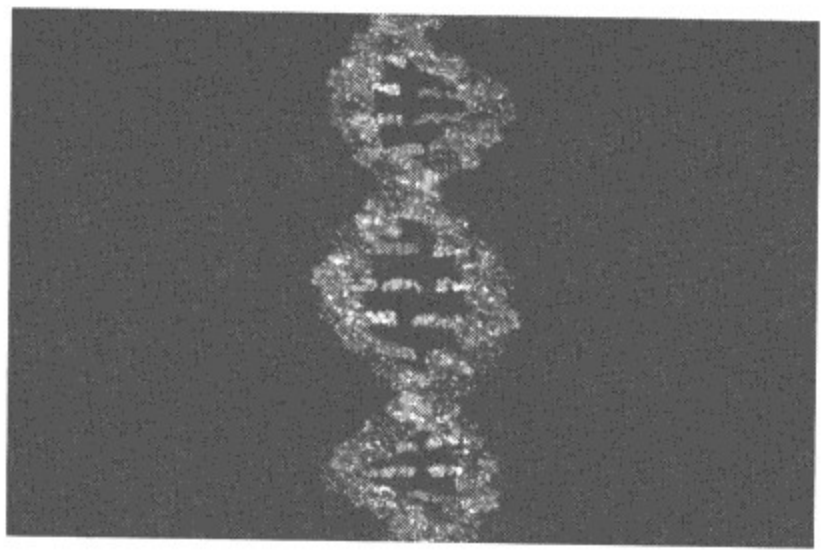
大多数基因是由排列在双螺旋上的双链 DNA 组成。其中一个链与另一个链互补，这意味着忠实的基因复制需要两个链分离并获得互补搭档。典型基因中碱基的线性序列携带合成蛋白质的信息。

产生基因产物的过程叫做基因表达，在转录和翻译两个步骤中发生。在转录中，RNA 聚合酶合成信使 RNA，它是基因遗传信息的拷贝。在翻译中核糖体“阅读”mRNA 并根据其指令合成蛋白质。因此，基因序列的一个改变（突变）可能引起蛋白质产物中的相应改变。

所有生物可划归为 3 个领域：细菌、真核生物和古菌。尽管表面上古菌与细菌类似，但其分子生物学特性更相似于真核生物。

翻译　马　纪　校译　张富春

第2章　基因的分子特性



DNA 结构的计算机模拟。© *Douglas Struthers/Stone/Getty.*

在开始详细研究基因的结构与活性以及这些概念背后的实验证据之前，需要有一个完整的轮廓。因此在第 2 章、第 3 章我们将丰富第 1 章中的分子生物学发展简史。本章开篇将首先讨论基因的分子行为。

2.1　遗传物质的特性

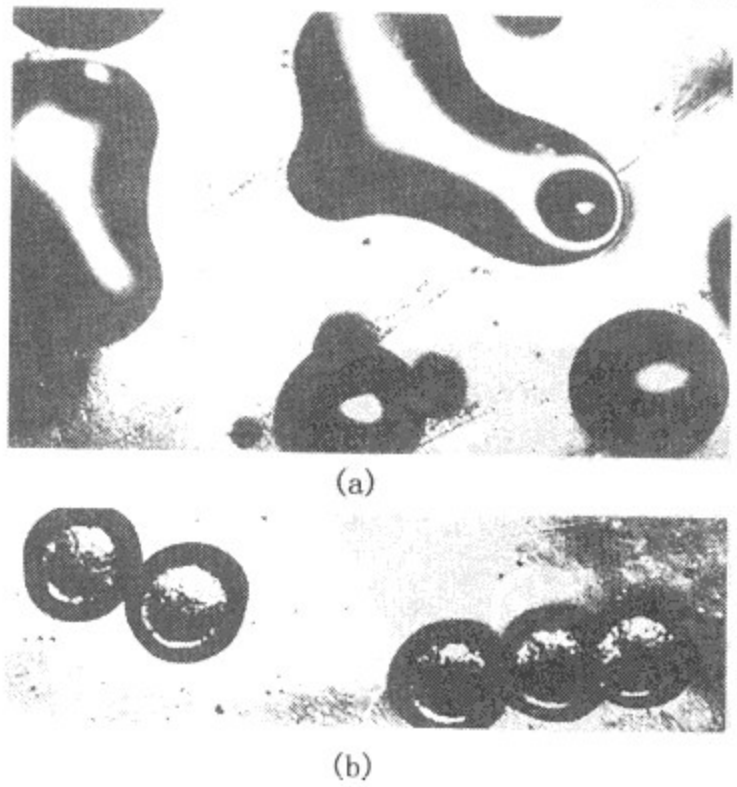
最终揭示基因化学本质的研究始于 1869 年德国的图宾根。Friedrich Miescher 从外科手术绷带的脓细胞（白细胞）中分离出细胞核（nuclei）。他发现细胞核中有一种新的含磷物质，并将其命名为核素（nuclein）。核素大部分是**染色质**（chromatin），一种脱氧核糖核酸（DNA）与染色体蛋白的混合物。

到 19 世纪末，从细胞的核酸与蛋白质混合物中分离出 DNA 和 RNA，从而有可能对核酸进行详细的化学分析（nucleic acid 一词及其缩写 DNA 和 RNA，直接源于 Miescher 的 nuclein）。从 19 世纪 30 年代开始，P. Levene 和 W. Jacobs 等已经证明 RNA 由 1 个糖（核糖）加上 4 个含氮碱基组成，而 DNA 则含有 1 个不同的糖（脱氧核糖）和 4 个碱基。他们发现每个碱基偶联 1 个糖-磷酸盐形成 1 个核苷酸（nucleotide）。在本章稍后部分我们再讨论 DNA 和 RNA 的化学结构。首先，让我们了解证明基因是由 DNA（有时是 RNA）组成的证据。

细菌转化

1928 年 Frederick Griffith 以其肺炎双球菌（*Streptococcus pneumoniae*）**转化**（transformation）实验奠定了 DNA 是遗传物质的基础。野生型肺炎双球菌是一种球形细胞，由叫做荚膜的黏液外壳所包裹，细胞形成大而亮的菌落，特征为光滑型（S）（图 2.1a）。这些细胞是有毒的（virulent）——注射到小鼠体内可引起致命性感染。*S. pneumoniae* 的某种突变株丧失了形成荚膜的能力，形成小而粗糙的（R）菌落（图 2.1b），是**无毒的**（avirulent）。由于没有保护性外壳，在它繁殖到足够产生任何危害之前，就被宿主的白细胞所吞噬了。

Griffith 的关键发现是热杀死的肺炎双球菌毒性



(a)

(b)

图 2.1　肺炎双球菌的变体。（a）大而亮的菌落光滑（S）且具毒性；（b）小而有斑点的菌落粗糙（R）且无毒性。（*Source*：（a，b）Harriet Ephrussi-Taylor.）

菌落能够将无毒细胞**转化**（transform）成毒性细胞。热杀死的毒性菌或活的无毒菌自身都不能引起致命性感染，但是混合到一起却是致死性的。这表明毒性特征以某种方式从死细胞传递到活的、无毒细胞。图 2.2 说明了这种转化现象。转化不是瞬时的，一旦无毒菌被赋予产荚膜杀死宿主的能力，就将其作为遗传特性传给后代。换句话说，在转化时无毒细胞通过某种方式从毒性细胞获得了基因。这意味着死菌中的转化物质可能是毒性基因。接下来的问题是转化物的化学本质是什么？

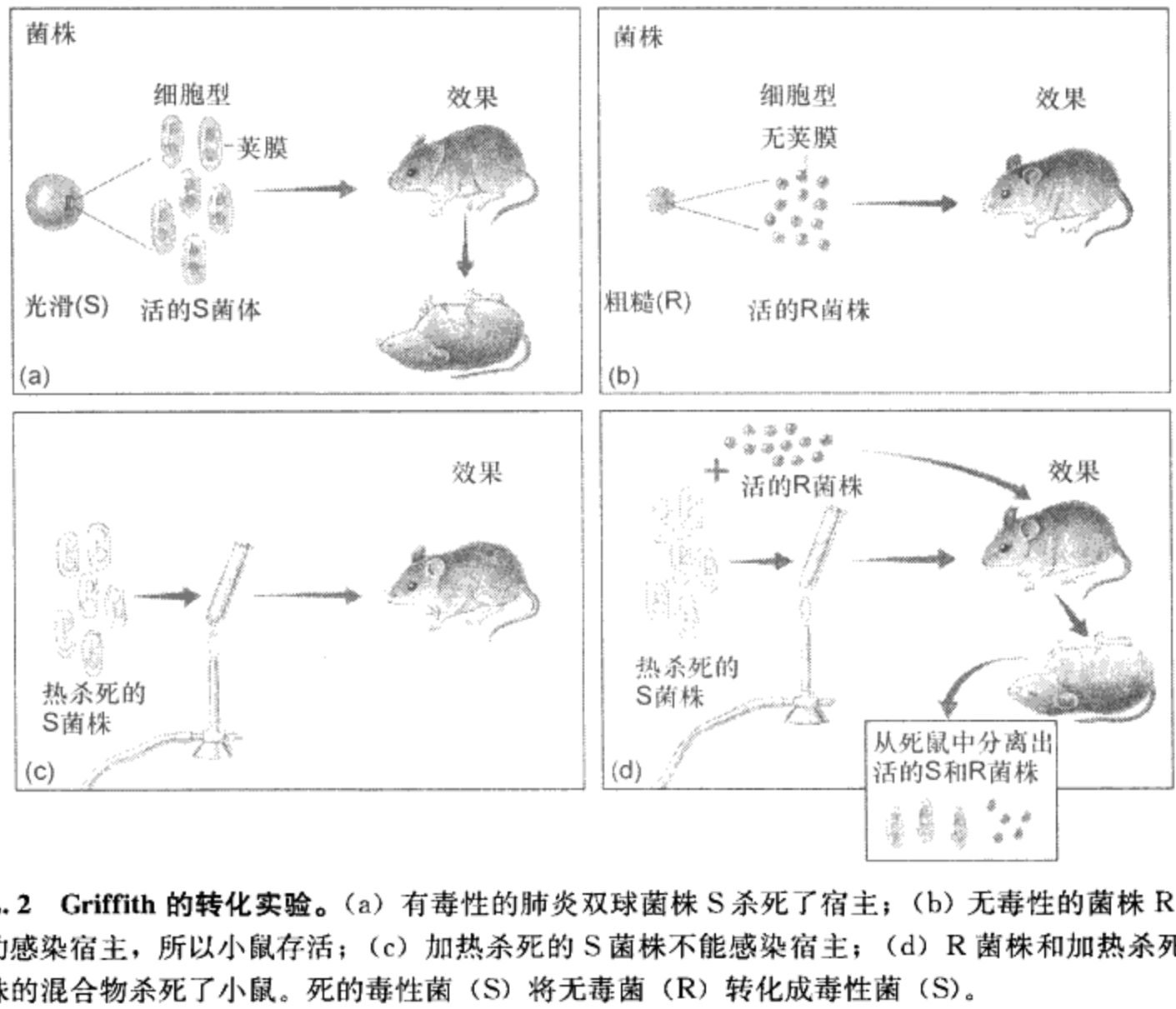


图 2.2　Griffith 的转化实验。（a）有毒性的肺炎双球菌株 S 杀死了宿主；（b）无毒性的菌株 R 不能成功感染宿主，所以小鼠存活；（c）加热杀死的 S 菌株不能感染宿主；（d）R 菌株和加热杀死的 S 菌株的混合物杀死了小鼠。死的毒性菌（S）将无毒菌（R）转化成毒性菌（S）。

DNA 是转化物质　1944 年，Oswald Avery、Colin MacLeod 和 Maclyn McCarty 采用与 Griffith 相似的转化实验证明了来自毒性细胞转化物质的化学本质。首先，用有机溶剂去除提取物中的蛋白质后，发现提取物仍然能转化。接下来，用各种酶消化提取物。降解蛋白质的胰蛋白酶和糜蛋白酶对转化无影响，降解 RNA 的核糖核酸酶也没有影响。这些实验排除了蛋白质和 RNA 是转化物。另外，Avery 及同事发现分解 DNA 的脱氧核糖核酸酶（DNase）破坏了毒性细胞提取物的转化能力，这些结果提示 DNA 是转化物质。

直接的物理化学分析结果支持了纯化的转化物是 DNA 的假设。Avery 及同事使用了如下分析工具。

1. *超速离心法*　在超速离心机中离心转化物来估计其大小。具有转化活性的物质很快沉降（快速移向离心管底），提示其具有很高的分子质量，即 DNA 的特征。

2. *电泳*　将转化物放入电场中观察其移动速率。转化的活性物具有相对较高的迁移率——也是 DNA 的特点，因为它有高的电荷/质量比。

3. *紫外吸收分光光度技术*　将转化物溶液置于分光光度计中，看哪种紫外线吸收最强。吸收光谱与 DNA 的吸收光谱匹配，即最强的吸收波长是 260nm，而蛋白质的最大吸收波长为 280nm。

4. *元素的化学分析*　得到氮/磷比均值为 1.67，正是富含这两种元素的 DNA 的预计值，远远低于蛋白质的预计值，蛋白质氮多磷少，微量蛋白质的污染都会提高氮/磷比例。

进一步确认　这些发现应该解决基因的特性了，但效果甚微。源自早期化学分析的错误概念认为 DNA 是一个四核苷酸的单一重复序列，如 ACTG-ACTG-ACTG 等，这使许多遗传学家认为 DNA 不可能是遗传物质。而且，还存在转化物中可能有蛋

白质污染，以及转化是否由控制 R 和 S 之外的其他基因完成，甚至细菌的基因与高等生物的基因是否一样等争论。

然而，到 1953 年沃森和克里克发表 DNA 双螺旋模型时，绝大多数遗传学家都已接受了基因由 DNA 构成。发生了什么变化呢？一件事是 1950 年 Erwin Chargaff 研究表明，在 DNA 中没有像以前的结果所提示的那样真正地发现碱基等比例，DNA 的碱基组成在物种之间不同。实际上，这正是对基因所期望的，基因在物种之间不同。此外，Rollin Hotchkiss 精练和扩展了 Avery 的发现。他将转化物纯化到只含有 0.02%的蛋白质，但仍能改变细菌细胞的遗传特性，进一步表明这种高纯度的 DNA 能够传递 R 和 S 之外的其他遗传性状。

1952 年 A. D. Hershey 和 Martha Chase 最终以另外一个实验进一步增加了基因由 DNA 组成的证据。这个实验用的是能感染大肠杆菌的 T2 **噬菌体**（图 2.3）。在感染期间，噬菌体基因进入宿主细胞，指导合成新的噬菌体颗粒。噬菌体仅由蛋白质和 DNA 组成。问题是基因存在于蛋白质中还是 DNA 中呢？Hershey-Chase 的实验回答了这个问题。感染时，大部分的 DNA 连同极少量蛋白质进入细菌中，大量的蛋白质外壳留在外面（图 2.4）。由于 DNA 是进入宿主细胞的主要成分，因此它可能含有基因。当然，这个结论还不具有说服力，随着 DNA 一起进入细胞的那些少量蛋白质也可能携带基因。但是将此前的结果放在一起分析，这个研究有助于使遗传学家相信遗传物质是 DNA 而非蛋白质。

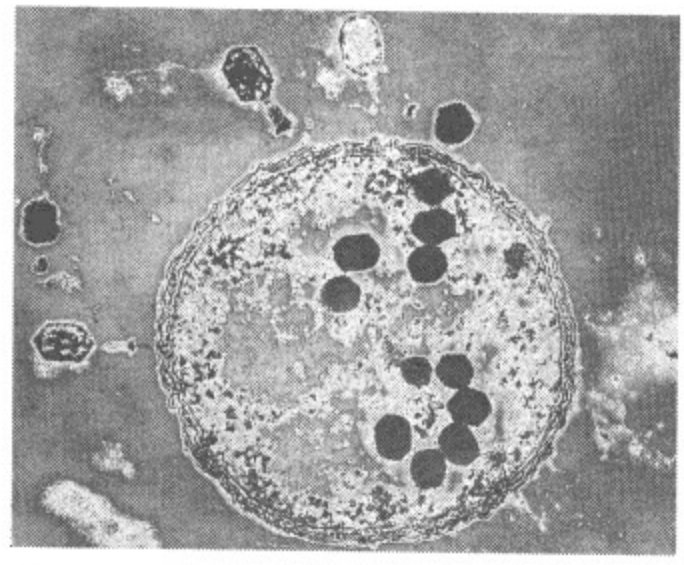


图 2.3　T2 噬菌体感染大肠杆菌细胞的透射电镜伪色照片。左上方噬菌体颗粒准备将其 DNA 注入宿主细胞内。另一个 T2 噬菌体已经感染了细胞，并且子代噬菌体颗粒正在组装，已经可以看出子代噬菌体的头部在宿主细胞内呈黑色多角形。（*Source*: © Lee Simon/Photo Researchers, Inc.）

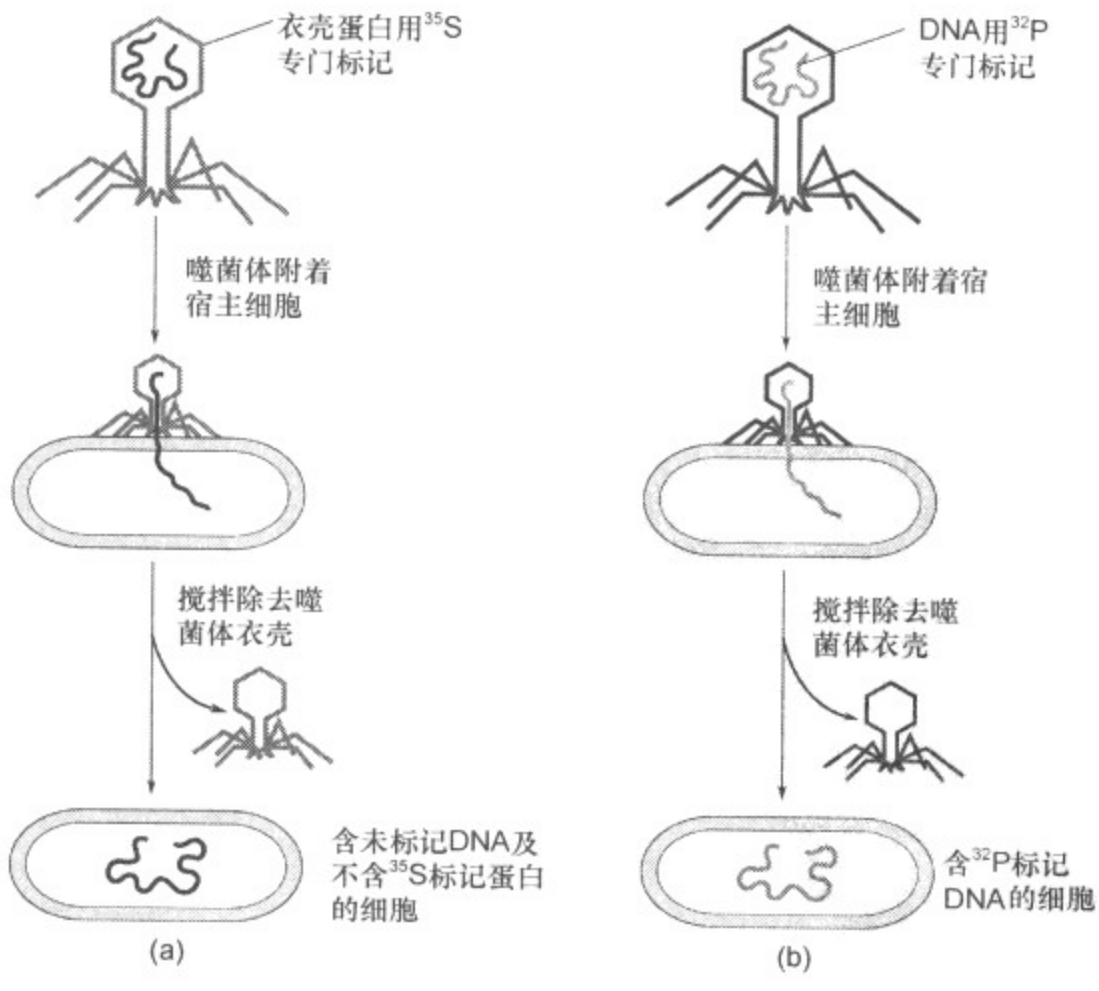


图 2.4　Hershey-Chase 实验。T2 噬菌体含有的基因使它能够在大肠杆菌中复制。由于噬菌体仅仅由 DNA 和蛋白质组成，它的基因肯定是由这些物质中的一种构成的。为了发现它的基因是由哪种物质组成的，Hershey 和 Chase 做了两部分实验。在第一部分中（a），用硫-35（^{35}S）（红色）标记噬菌体的蛋白质，DNA 不标记（黑色）。在第二部分中（b），用磷-32（^{32}P）（橘红色）标记噬菌体 DNA，蛋白质不标记（黑色）。由于噬菌体基因必须进入细胞，实验者通过对感染细胞的标记类型进行推断揭示基因的特性。大部分标记的蛋白质留在外面，通过搅拌使标记的蛋白质脱离细胞（a），而大部分标记的 DNA 进入感染的细胞（b）。结论是该噬菌体的基因由 DNA 构成。

Hershey-Chase 实验取决于 DNA 和蛋白质上不同的放射性同位素标记。用磷-32（^{32}P）标记 DNA，用硫-35（^{35}S）标记蛋白质。这样选择是因为 DNA 富含磷而噬菌体蛋白不含磷，蛋白质含硫而 DNA 不含硫。

Hershey 和 Chase 将标记过的噬菌体通过尾部吸附到细菌上，并将其基因注入宿主。接着，用搅拌器剧烈混合除去噬菌体的空外壳。他们知道基因肯定进入了细胞，问题是什么进去了？是^{32}P标记的 DNA 还是^{35}S标记的蛋白质？正如我们已经看到的，是 DNA！一般来说，基因是由 DNA 组成的。在本章后面将看到，其他实验显示某些病毒基因是由 RNA 组成的。

小结 基因由核酸组成，通常是 DNA。一些简单的遗传系统（如病毒的基因）是 RNA。

多聚核苷酸的化学本质

到 20 世纪 40 年代中期，生物化学家知道了 DNA 和 RNA 的基本化学结构。把 DNA 分解成相应的组成部分时，发现这些组分含**氮碱基**（base）、**磷酸**（phosphoric acid）和**脱氧核糖**（deoxyribose）（故命名为脱氧核糖核酸）。同样地，RNA 由碱基、磷酸和一种不同的糖［**核糖**（ribose）］组成。在 DNA 中发现的 4 种碱基是腺嘌呤（A）、胞嘧啶（C）、鸟嘌呤（G）和胸腺嘧啶（T）。RNA 中的碱基除了尿嘧啶（U）代替胸腺嘧啶（T）外，其他碱基相同，这些碱基结构如图 2.5 所示，腺嘌呤和鸟嘌呤与母体分子——嘌呤有关。因此，我们称这些化合物为**嘌呤**（purine）。另外两个碱基像嘧啶，被称为**嘧啶**（pyrimidine）。这些结构组成了遗传学的基础。

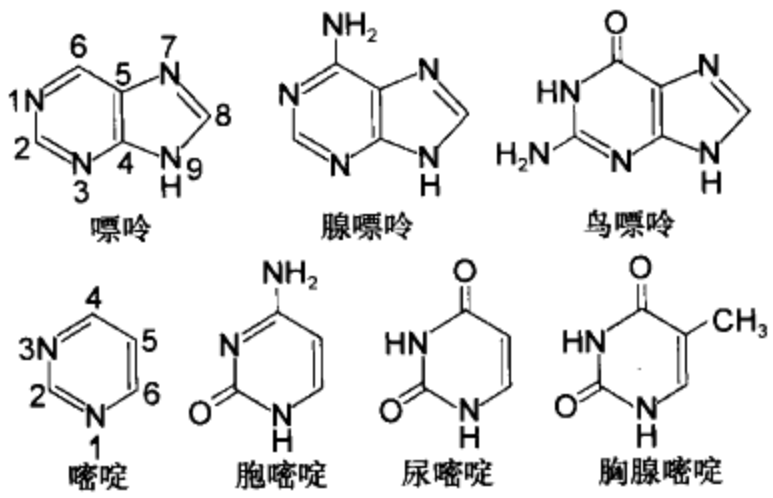


图 2.5 DNA 和 RNA 的碱基。左边的嘌呤和嘧啶在 DNA 和 RNA 中不会出现，只是用来与其他 5 个碱基比较。

图 2.6 描绘了核酸中糖的结构。注意，它们只在 2′位置有区别。核糖有一个羟基，而脱氧核糖缺少氧只有一个氢（H），用竖线表示，因此命名为脱氧核糖。RNA 和 DNA 中的碱基和糖连接在一起形成的单位称之为**核苷**（nucleoside）（图 2.7）。核苷的名称来源于相应的碱基：

碱基	核苷（RNA）	脱氧核苷（DNA）
腺嘌呤	腺嘌呤核苷	脱氧腺嘌呤核苷
鸟嘌呤	鸟嘌呤核苷	脱氧鸟嘌呤核苷
胞嘧啶	胞嘧啶核苷	脱氧胞嘧啶核苷
尿嘧啶	尿嘧啶核苷	不常见
胸腺嘧啶	不常见	（脱氧）胸腺嘧啶核苷

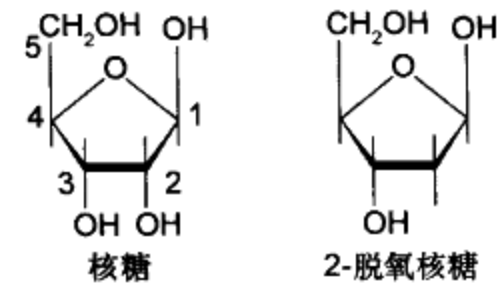


图 2.6 核酸的糖。注意在核糖第 2 位的羟基及其在脱氧核糖第 2 位的缺失。

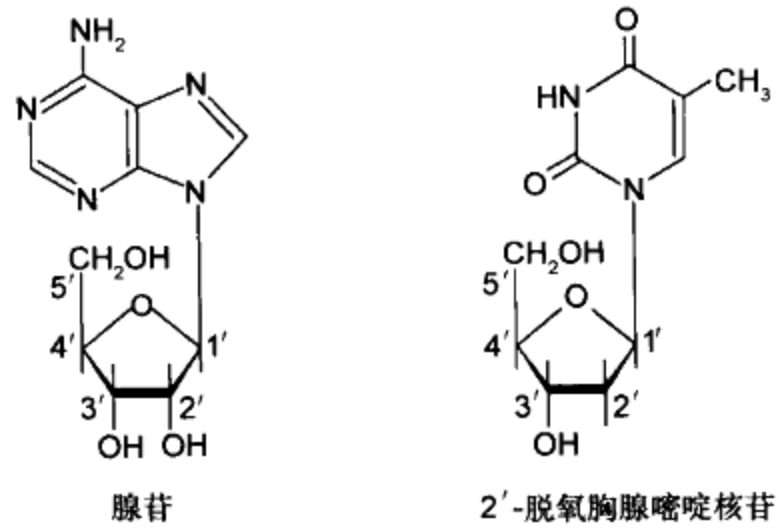


图 2.7 核苷的两个例子。

由于胸腺嘧啶在 RNA 中不常见，其核苷“脱氧”名称经常是假定的。因此其脱氧核苷酸简称胸腺嘧啶核苷（thymidine）。在核苷中，糖的碳原子编号（图 2.7）很重要。在碱基中用普通数字标注，在糖中用带“′”的数字标注。例如，碱基连接在糖的 1′位。在脱氧核苷中 2′位是脱氧的。DNA 和 RNA 中的糖通过其 3′和 5′位连接。

图 2.5 中的结构采用有机化学的速写，为简洁起见，省去某些原子。图 2.6 和图 2.7 则稍有不同，其中带自由端的直线表示末端有氢原子的 C—H 键。图 2.8 是腺嘌呤和脱氧核糖结构的速写和包括每个原子的详图。

DNA 和 RNA 的基本组成单位是**核苷酸**（nucleotide），由核苷通过磷酯键与磷酸基团连接（图 2.9）。酯键是由醇（含羟基）与酸形成的有机复合物。在核苷酸里，醇基是糖的 5′羟基，酸是磷酸，所

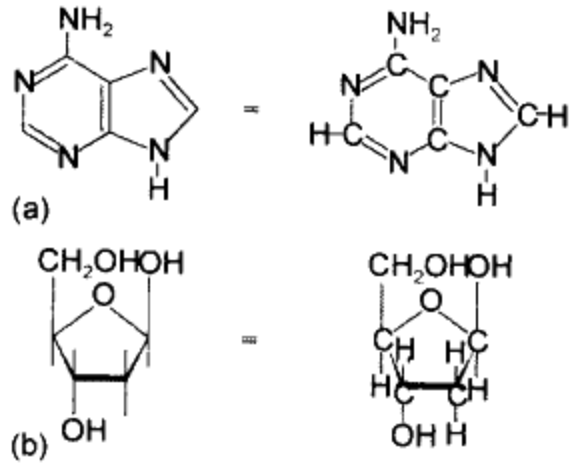


图 2.8　腺嘌呤和脱氧核糖的结构。注意左边的结构没有标出大部分或全部碳原子及某些氢原子。在右边的结构中分别用红色和蓝色标出。

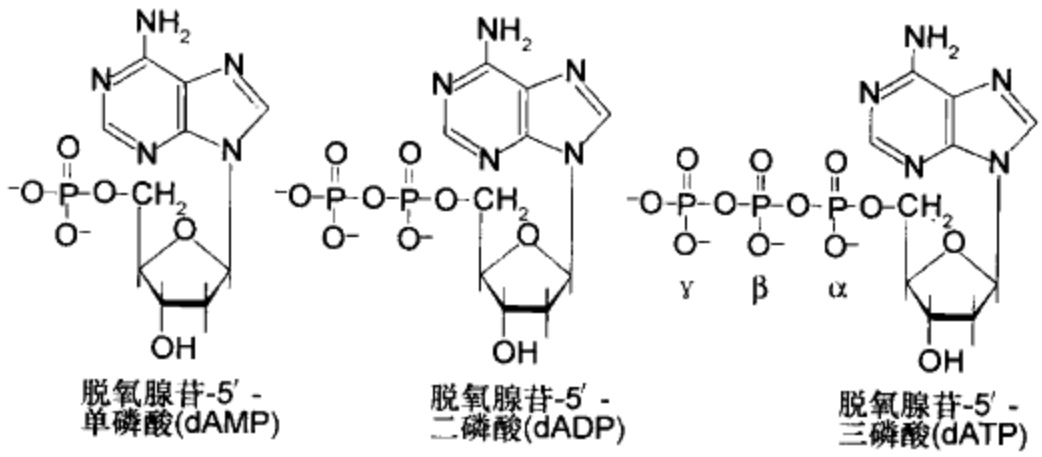


图 2.9　3 种核苷酸。脱氧腺苷的 5′-核苷酸通过 5′羟基磷酸化形成，加上一个磷酸形成脱氧腺苷-5′-磷酸（dAMP）。再加一个磷酸形成脱氧腺苷-5′-二磷酸（dADP）。3 个磷酸（标记为 γ、β、α）形成脱氧腺苷-5′-三磷酸（dATP）。

以称其为磷酸酯键。图 2.9 是 DNA 4 个前体之一的脱氧腺苷-5′-三磷酸（dATP）的结构。在 DNA 合成时，2 个磷酸基从 dATP 中移走，剩下脱氧腺苷-5′-磷酸（dAMP）。DNA 的其他 3 个核苷（dCMP、dGMP 和 dTMP）有类似的结构和名称。

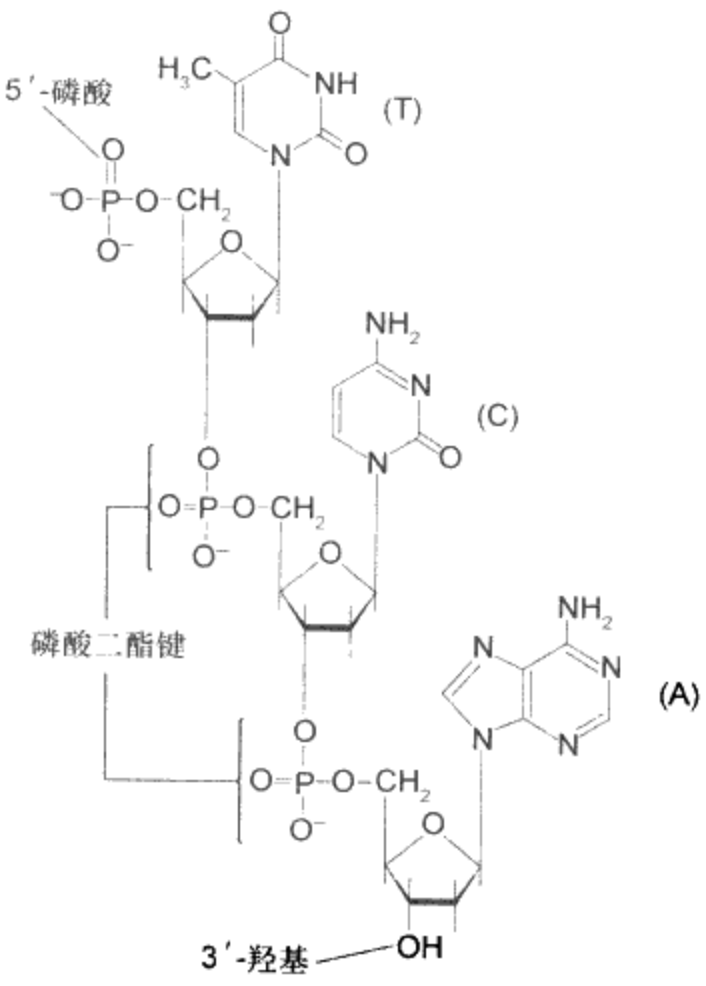


图 2.10　三核苷酸。这个小 DNA 片段只含有 3 个核苷酸，它们通过糖的 5′和 3′羟基之间的磷酸二酯键（红色）相互连接。这个 DNA 的 5′端在顶部，有一个自由的 5′磷酸基；3′端在底部，有一个自由的 3′羟基。这个 DNA 的序列可读成 5′pdTpdCpdA3′，通常被简写成 TCA。

我们将在第 20 章和第 21 章详细讨论 DNA 的合成。现在请注意在 DNA 和 RNA 中把核苷酸连接在一起的键的结构（图 2.10）。这些键叫做**磷酸二酯键**（phosphodiester bond），使磷酸与两个糖连接在一起，一个与糖的 5′基团相连，另一个与 3′基团相连。你会注意到，相对于前面图中的位置，这张图中碱基被旋转了，这与它们在 DNA 和 RNA 中的几何关系更接近。还要注意这个**三核苷酸**（trinucleotide），或者 3 个核苷酸串，具有极性：分子顶部有一个游离的 5′磷酸基，称为 5′端（5′-end），底部有一个游离的 3′羟基，称为 3′端（3′-end）。

图 2.11 介绍一种表示核苷酸和 DNA 链的速写方法。用垂直线表示脱氧核糖，碱基连接在顶端的 1′位，磷酸二酯通过 3′（中间）和 5′（底部）位连接相邻的核苷酸。

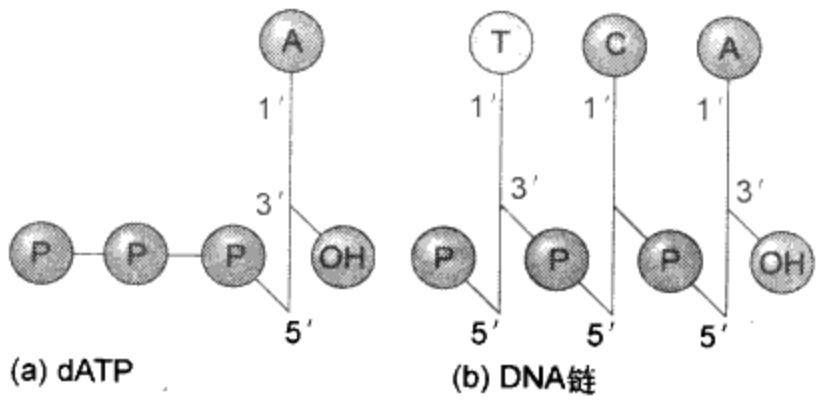


图 2.11　DNA 的速写注释。（a）核苷酸 dATP。突出 DNA 组成模块的 4 个特征：①脱氧核糖用垂直黑线表示；②糖的 1′位置连接的是腺嘌呤碱基（A）；③糖的 3′位连接一个羟基（OH）；④糖的 5′位连接一个三磷酸基团。（b）一条短的 DNA 链。在图 2.10 中的相同三核苷酸在这里用速写法表示。注意 5′端磷酸、磷酸二酯键和 3′端羟基（—OH）。习惯上这小段 DNA 从左到右、从 5′端到 3′端开始写。

小结 DNA 和 RNA 是链状分子，由称之为核苷酸的亚基组成。核苷酸包含一个碱基和一个磷酸基团，碱基连接在糖（在 RNA 中为核糖，在 DNA 中为脱氧核糖）的 1′位置。磷酸通过磷酸二酯键与 DNA 或 RNA 的糖基连接。

2.2 DNA 结构

以上所介绍的所有关于 DNA 和 RNA 的知识是到 20 世纪 40 年代末才获知的。那时 DNA 是遗传物质也变得清楚了，因此它处于生命研究的最核心位置。然而，DNA 的三维空间结构还未知，许多研究者致力于发现该结构。

实验背景

加州理工学院的理论化学家 Linus Pauling 是众多对 DNA 结构感兴趣的科学家之一，在化学键及蛋白质 α 螺旋结构的研究方面闻名遐迩。α 螺旋是靠氢键聚合在一起的，这为沃森和克里克提出 DNA 双螺旋模型提供了知识基础。另一个研究小组是伦敦国王大学的 Maurice Wilkins、Rosalind Franklin 及同事。他们用 X 射线衍射分析 DNA 的空间三维结构。最后，沃森和克里克也加入到这个竞赛中。沃森当时只有 20 出头，但已获得印第安纳大学的博士学位，在英国剑桥大学卡文迪许实验室研究 DNA。在那里他遇见了 35 岁的物理学家克里克，当时克里克正参加转为分子生物学家的培训。沃森和克里克没有亲自做实验，他们的策略是利用其他研究小组的数据构建 DNA 模型。

Erwin Chargaff 是另一位非常重要的贡献者。我们已经看到他 1950 年的论文是怎样证明 DNA 是遗传物质的，但是论文中还含有另一条更有意义的信息。Chargaff 对不同来源 DNA 碱基组成的研究发现，嘌呤的含量总是大体与嘧啶的含量相等。此外，腺嘌呤与胞嘧啶的量总是大体相等，鸟嘌呤与胸腺嘧啶也是如此。这就是 Chargaff 法则（Chargaff's rule），这一法则为沃森和克里克模型奠定了宝贵的基础。表 2.1 展示了 Chargaff 的数据。你会注意到由于一些碱基的回收不完全，略偏离法则，但整体模式是明了的。

或许揭开 DNA 结构之谜最关键的一幕是来自 Franklin 在 1952 年拍摄的一张 DNA X 射线衍射照片——那张 Wilkins 于 1953 年 1 月 30 日在伦敦给沃森看的照片。X 射线技术操作如下：实验者制备非常浓而黏的 DNA 溶液，然后用一根针挑出 DNA 拉成纤维状。这根纤维不是单个的分子，而是一整束 DNA 分子，通过从溶液中拽出迫使它们并肩整齐排列。如果周围空气中的相对湿度恰当，这根纤维足以像晶体那样以可解释的方式衍射 X 射线。实际上在 Franklin 的照片中（图 2.12），衍射图非常简单，只是一系列的点排列成 X 形状，表明 DNA 结构本身一定非常简单。相反，像蛋白质那样复杂不规则的分子产生的 X 射线衍射图很复杂，有许多斑点，更像用散弹猎枪射击后的表面。由于 DNA 分子非常大，只有具备规则、重复的结构，它才可能有简单的 X 射线衍射照片。而对一条细长分子所能推测的最简单重复形状是螺旋状的瓶塞钻样（cork-screw）的结构或螺旋（helix）。

双螺旋

Franklin 的 X 射线研究有力地显示 DNA 是一种螺旋。不仅如此，它还提供了有关螺旋形状和大小的重要信息。具体地讲，在×图形一条臂上相邻条带的间隔与螺旋中总的重复距离 33.2Å 反向相关，×图形顶端与底部间的距离与螺旋上重复元件（**碱基对**，base pair）之间的距离（3.32Å）反向相关（见第 9 章 Bragg 规则解释这种反向关系）。虽然

表 2.1 每摩尔磷酸中 DNA 的摩尔碱基组成

	人类				酵母		禽结核杆菌	牛				
	精子		胸腺	肝癌				胸腺			脾	
	#1	#2			#1	#2		#1	#2	#3	#1	#2
A:	0.29	0.27	0.28	0.27	0.24	0.30	0.12	0.26	0.28	0.30	0.25	0.26
T:	0.31	0.30	0.28	0.27	0.25	0.29	0.11	0.25	0.24	0.25	0.24	0.24
G:	0.18	0.17	0.19	0.18	0.14	0.18	0.28	0.21	0.24	0.22	0.20	0.21
C:	0.18	0.18	0.16	0.15	0.13	0.15	0.26	0.16	0.18	0.17	0.15	0.17
回收率	0.96	0.92	0.91	0.87	0.76	0.92	0.77	0.88	0.94	0.94	0.84	0.88

Source: E. Chargaff "Chemical Specificity of Nucleic Acids and Mechanism of Their Enzymatic Degradation", *Experientia* 6: 206, 1950.

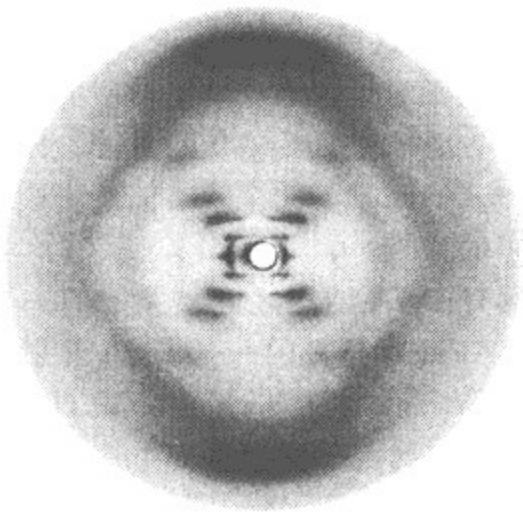


图 2.12　Franklin 的 DNA X 射线衍射图。图像的规则性图案表明 DNA 是一个螺旋。×的顶端和底部带间的间距给出螺旋元件（碱基对）间的间距为 3.4Å。图案中相邻条带间的间隔螺旋的全部重复（一个螺距的长度）为 34Å。（*Source*: Courtesy Professor M. H. F. Wllkins, Biophysics Dept., King's College, London.）

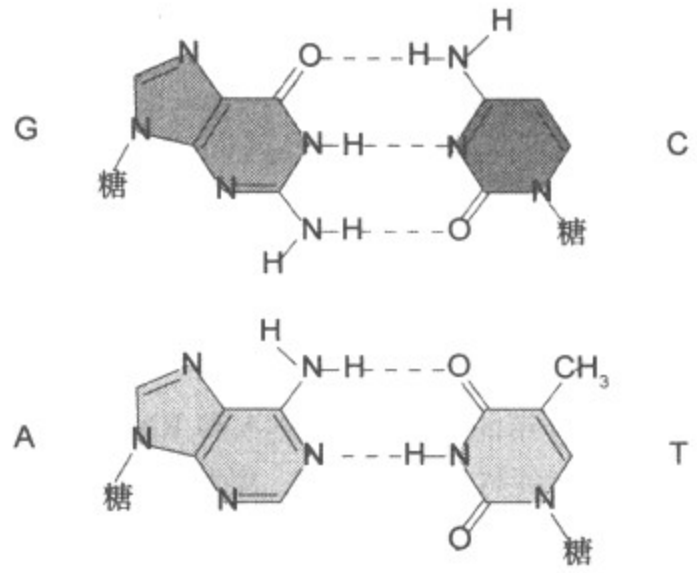


图 2.13　DNA 的碱基对。通过 3 个氢键连接在一起的 G-C 碱基对（虚线），与通过 2 个氢键连接在一起的 A-T 碱基对形状几乎完全相同。

Franklin 的衍射图说明了 DNA 的很多东西，但也带来一个悖论：DNA 是规则的、有重复结构的螺旋，而它要行使遗传功能，就应该是无规则的碱基顺序。

沃森和克里克想出了一种模式，既能解决这个矛盾同时又能满足 Chargaff 定律，即 DNA 必须是**双螺旋**（double helix），糖-磷酸骨架在外面，碱基在里面。此外，碱基必须配对，一条链上的嘌呤总是与另一条链上的嘧啶配对。这种方式的螺旋是均一的，不会产生两个大的嘌呤配对的膨胀区或两个小的嘧啶配对的收缩区。沃森开玩笑地讲述他想到双螺旋的原因说："我已经决定建立双链模型。Francis 不得不同意。虽然他是物理学家，但也知道重要的生物学目标是成对出现的。"

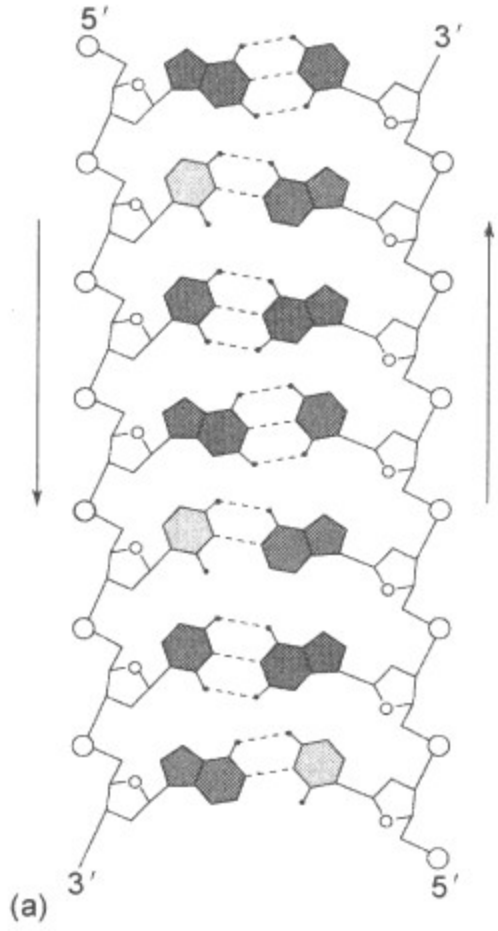
但是 Chargaff 规则远不止这些，由其判断腺嘌呤与胸腺嘧啶等量，同样鸟嘌呤和胞嘧啶也是等量的，这与沃森和克里克的观测结果非常吻合，通过氢键结合的 A-T 碱基对的形状与 G-C 碱基对几乎完全相同（图 2.13）。所以，沃森和克里克推测腺嘌呤必须总是与胸腺嘧啶配对，鸟嘌呤必须总是与胞嘧啶配对。这样，双链 DNA 将是均匀的，由形状非常相似的碱基对组成，不必考虑任一 DNA 单链本身的不可预知序列。这是他们至关重要的洞察力，是 DNA 结构的关键。

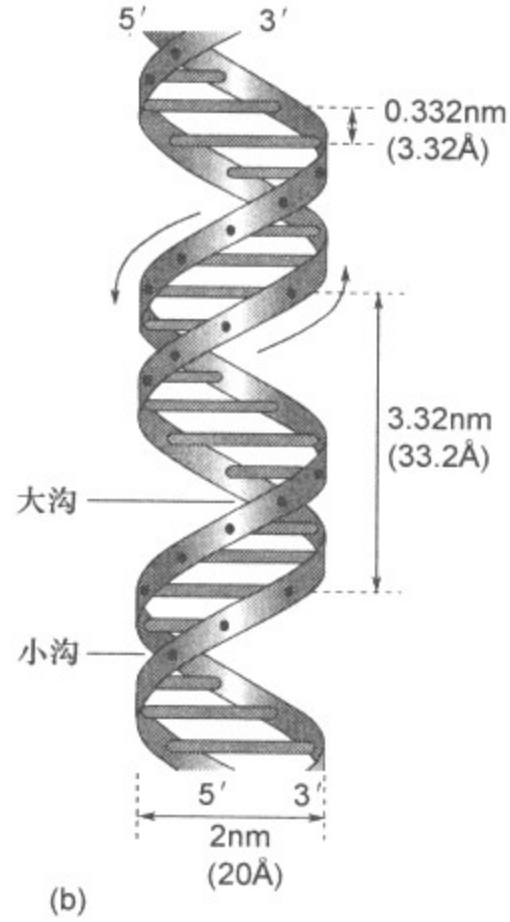
双链螺旋经常被画成扭曲的梯子，在图 2.14 中以 3 种方法表示。梯子弯曲的两边代表两条 DNA 链的糖-磷酸骨架，横档代表碱基对。碱基对之间的距离是 3.32Å，全部螺旋重复距离（螺距）是 33.2Å，意味着螺旋每旋一周有 10 个碱基对（bp）。箭头表示这两条链是**反向平行的**（antiparrallel）。如果一条链从上到下具有 5′→3′极性，那么另一条链从上到下必定具有 3′→5′极性。在溶液中，DNA 的结构与此非常类似，但螺旋每一圈约含 10.4 个碱基对。

沃森和克里克在 *Nature* 上发表了 DNA 模型的轮廓，背面是 Wilkins 和 Frankin 及同事关于 X 射线衍射数据的论文。沃森-克里克论文是简洁的经典——只有 900 个字，还不到一页。投稿之后不到一个月就迅速发表了。实际上，克里克想清楚地阐明模型的生物学含义，但沃森不乐意那样做。他们斟酌了一句在科学文献中最伟大的含糊句子，"我们已经注意到我们所提出的特殊碱基配对规则立刻会使人联想到遗传物质的某种可能的复制机制。"

正如这句引起争论的话所预示的，沃森-克里克模型确实提出了一种 DNA 复制机制。因为两条链互补（complement）可被分开，然后各自作为模板产生一条新的互补链。图 2.15 示意这一过程是怎样完成的。注意这种叫做半保留复制的机制怎样保证两条子代 DNA 链与母链完全相同，并保证细胞分裂时基因的整体性。1958 年 Matthew Meselson 和 Franklin Stahl 证明这确实是 DNA 的复制方式（见第 20 章）。

小结　DNA 分子是一种双链螺旋，糖-磷酸骨架位于螺旋的外面，碱基对位于螺旋的里面。碱基以特殊的方式配对：腺嘌呤（A）与胸腺嘧啶（T）配对，鸟嘌呤（G）与胞嘧啶（C）配对。





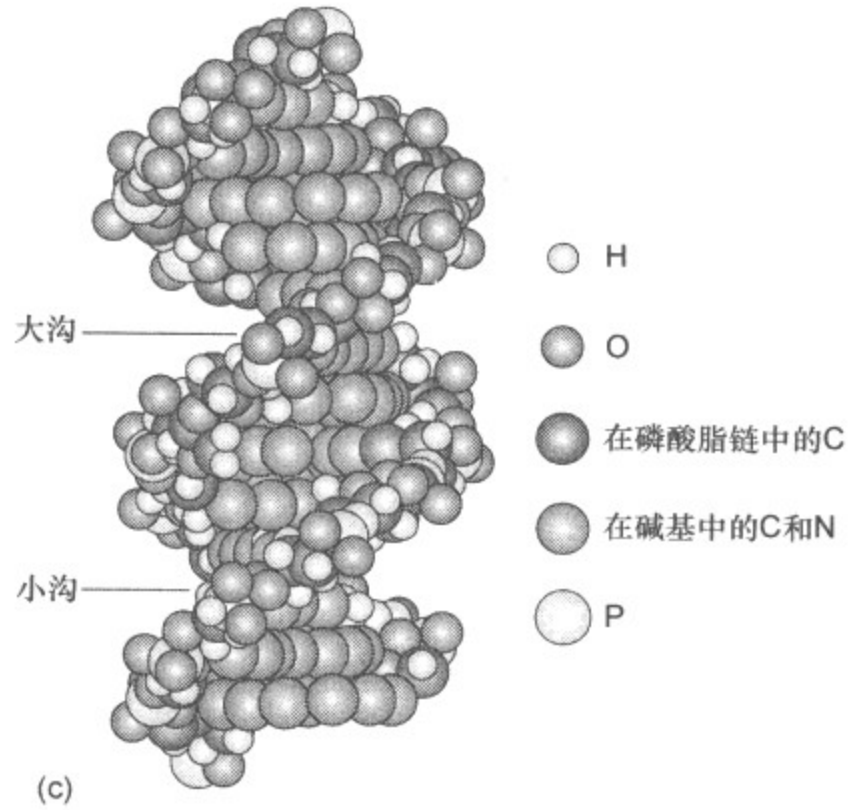


图 2.14　DNA 结构的 3 种模型。（a）螺旋拉直显示在中间的碱基配对。每种碱基用不同颜色表示，糖-磷酸骨架用黑色表示。注意 G-C 碱基对中有 3 个氢键，而 A-T 碱基对中有 2 个氢键。每一条链旁边的垂直箭头表示 5′→3′方向，两条 DNA 链反向平行。左边的链从高到低，由 5′→3′方向延伸；右边的链从低到高由 5′→3′方向延伸。脱氧核糖环（用 O 代表氧的白色五角形）同样显示两条链方向相反：右边链中的脱氧核糖环与左边的成反向关系。（b）双链螺旋像一个扭曲的梯子，梯子的两边代表两条 DNA 链的糖-磷酸骨架，横档代表碱基对。两条链旁边弯曲的箭头指出每条链从 5′→3′的方向，进一步阐明两条链反向平行。（c）空间填充模型。糖-磷酸骨架由黑色、红色、白色和黄色小球串起来，碱基对是蓝色小球组成的水平平面。注意大小沟在（b）和（c）指出。

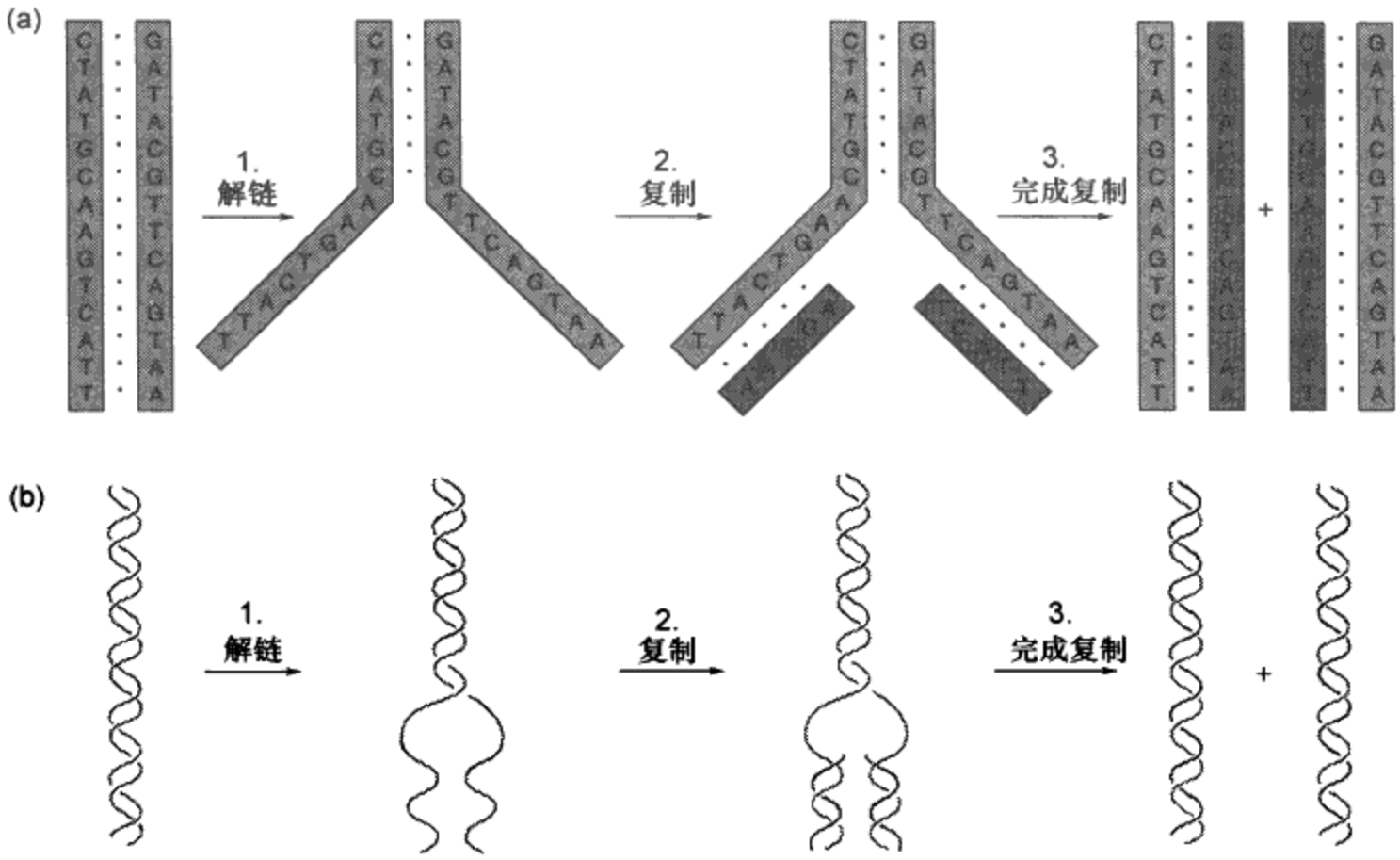
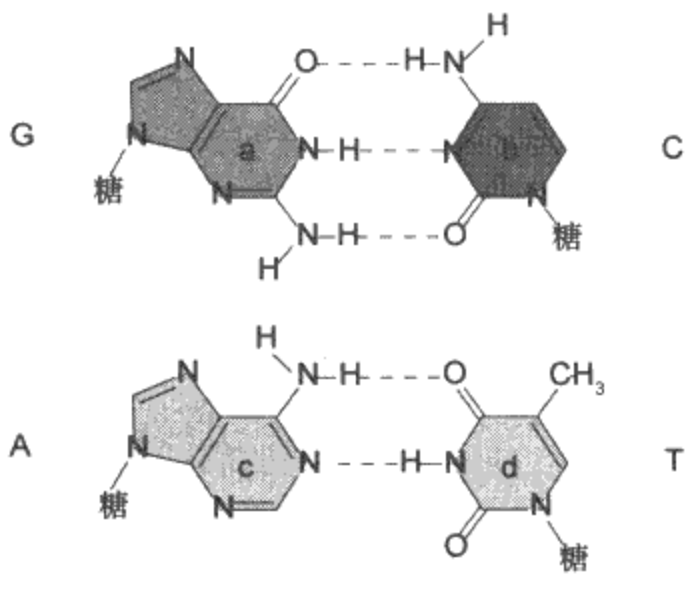


图 2.15　DNA 的复制。(a) 简化起见，两条 DNA 母链用平行线表示。步骤 1：复制时母链分离或解旋。步骤 2：与分开的母链互补的碱基合成新链。步骤 3：随着母链的完全分离及新链的完成，复制结束。最终结果是产生了与母链完全相同的两条 DNA 双链。因此，每个子双链有一条母链和一条新链。由于每个子双链只保留了一条母链，因此这种复制机制称之为半保留复制。(b) 相同过程的更形象描述。双链螺旋中显示的曲线代替了平行线。再次注意两个子双螺旋的产生，每个都有一条母链和一条新链。

问题解答

问题 1

下图是两个碱基对的轮廓，用 a、b、c 和 d 表示碱基。这些碱基的真实身份是什么？



解答

碱基 a 是嘌呤（由 1 个六环和 1 个五环组成），所以我们知道它一定是腺嘌呤（A）或鸟嘌呤（G）。而且，它通过 3 个氢键（虚线）连接其配对碱基。G-C 碱基对通过 3 个氢键连接，而 A-T 碱基对只有 2 个氢键连接。因此，碱基 a 是鸟嘌呤，配对碱基 b 是胞嘧啶。由于 A-T 碱基对是通过 2 个氢键连接的，碱基 c 一定是腺嘌呤（A），其配对碱基 d 必定是胸腺嘧啶（T）。

问题 2

以下是 DNA 片段一条链的序列：

5′ATGCGTGACTAATTCG3′

写出这条 DNA 双链形式的序列。

解答

遵循沃森-克里克碱基配对原则，记住两条 DNA 链是反向平行的，双链 DNA 如下：

5′ATGCGTGACTAATTCG3′

3′TACGCACTGATTAAGC5′

2.3　RNA 基因

Hershey 和 Chase 运用的遗传系统是噬菌体，一种细菌病毒。病毒颗粒本身其实是基因的包装。它没有自己的生命，没有代谢活性，是无活性的。

但是当它感染宿主细胞时，似乎就活了。宿主细胞突然间开始制造病毒蛋白，然后病毒基因被复制，新制造的基因与病毒外壳蛋白组装成子代病毒颗粒。由于在宿主体外是无活力颗粒，进入体内却像有生命的东西，所以病毒很难分类。有些科学家称其为“活物质”或者“生物体”（organism），有些科学家倾向用一个更能描述病毒低于生命状态（less-then-living status）的词——“遗传系统”。这个词适用于任何含有复制能力的基因实体——真正的生物体或者病毒。

目前为止所研究的大多数遗传系统的基因都是由DNA组成的。但是有些病毒，包括几种噬菌体、植物和动物病毒（如艾滋病病毒HIV），具有RNA基因。有时病毒RNA基因是双链的，但通常都是单链的。

我们已经遇到了在分子生物学研究中利用病毒的著名例子，在以后的章节将看到更多的例子。实际上，没有病毒，分子生物学领域会出现无法估计的乏味。

小结 某些病毒所含基因由RNA而非DNA组成。

2.4 核酸的物理化学性质

可以推测DNA和RNA分子呈现出几种不同的结构，让我们来看看这些结构以及在一些条件下促使两条链先分离再结合DNA的特性。

DNA结构的多样性

由沃森和克里克提出的DNA结构（图2.14）表示在非常高的相对湿度（92%）下所形成的纤维束中DNA钠盐的结构，被称为B型DNA。虽然它可能接近细胞中大多数DNA的构型，但并不是双链核酸的唯一构象。如果将DNA纤维周围的相对湿度降到75%，则DNA钠盐呈现A型（图2.16a）。它在许多方面不同于B型（图2.16b）。最明显的是碱基对的平面不再垂直于螺旋轴，而是偏离水平面20°。还有，A型螺旋每个螺距包含10.7个碱基对而非B型螺旋晶体结构中的10个碱基对，每个螺距只有24.6Å而非33.2Å。这意味着螺旋一圈的距离仅有24.6Å而不是B-DNA中的33.2Å。具有一条DNA链和一条RNA链的杂交多聚核苷酸在溶液中呈A型结构，与双链RNA构象相同。表2.2给出了A型和B型DNA，以及将在下一段讨论的左手螺旋Z型DNA的螺旋参数。

A型和B型DNA都是右手螺旋：无论从顶端还是底端看，螺旋都是顺时针方向。1979年Alexander Rich和同事发现DNA并不总是右手螺旋的，含有交替的嘌呤和嘧啶（如多聚［dG-dC］·多聚［dG-dC］）

—GCGCGCGC—
—CGCGCGCG—

的DNA双链可存在于延伸的左手螺旋中。由于从侧面看这种DNA的骨架呈Z字形，所以常被称为Z-DNA。图2.16c是Z-DNA结构，其螺旋参数见表2.2。尽管Z-DNA是Rich在研究模式化合物多聚［dG-dC］·［dG-dC］时发现的，这种结构似乎不只是一种实验室现象。有证据显示活细胞中含有少部分Z-DNA。此外，Keji Zhao及同事在2001年发现，至少一种基因的激活需要调节序列转变成Z型DNA。

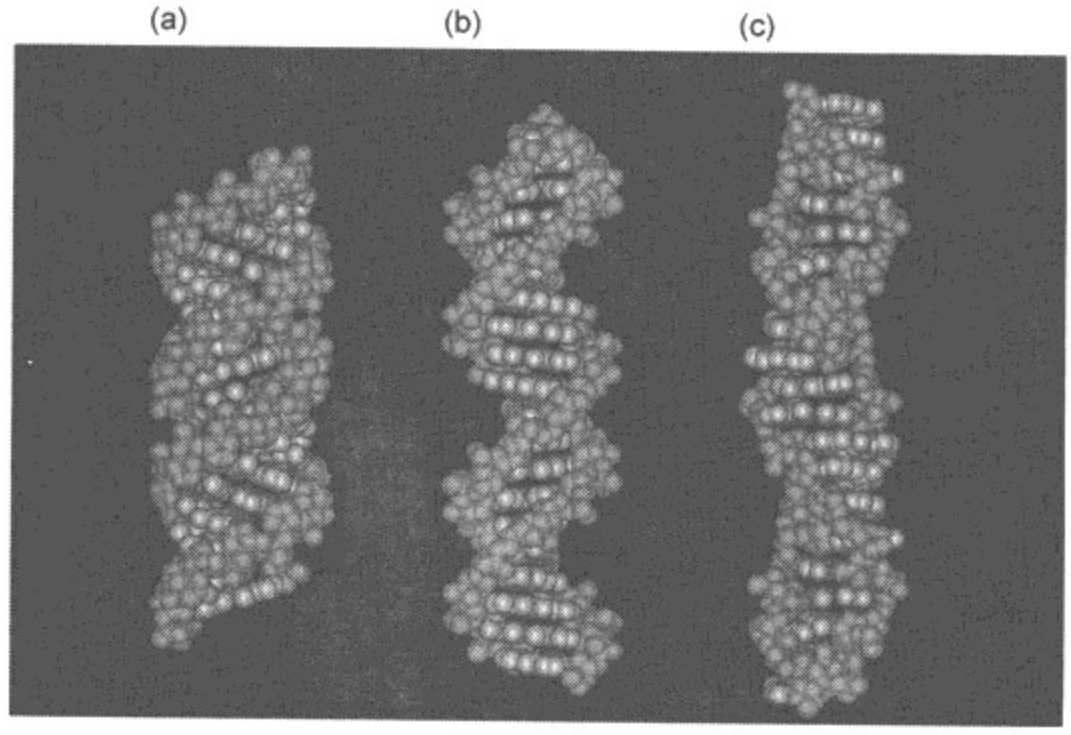


图2.16 A-DNA、B-DNA和Z-DNA的计算机模型。（a）A-DNA。注意其碱基对（蓝色）从右向左的倾斜在上部和底部附近的大沟中特别清楚。还请注意右手螺旋由糖-磷酸骨架（红色）显示。（b）B-DNA。注意熟悉的右手螺旋，大致水平的碱基对。（c）Z-DNA。注意是左手螺旋。这些DNA的碱基对数目都相同，强调3种结构在致密度上的不同。（*Source*：Courtesy Fusao Takusagawa.）

表 2.2　DNA 形式

形式	斜度/Å	每转碱基数	碱基对偏离水平面/°
A	24.6	10.7	+19
B	33.2	约 10	−1.2
Z	45.6	12	−9

小结　在细胞中，DNA 可能以常见的 B 构型存在，其碱基对是水平的。小部分 DNA 可能呈现伸展的称之为 Z-DNA 左手螺旋（至少在真核细胞中）。RNA-DNA 杂交链呈现出第三种螺旋，称之为 A 构型，其碱基对与水平面倾斜。

DNA 双螺旋两条链的分离　虽然一种生物的 DNA 中 G/C 和 A/T 的比例是固定的，但是在不同 DNA 中 GC 含量（G+C 的百分率）变化相当大。表 2.3 列出了几种生物和病毒 DNA 中 GC 的含量，变化范围为 22%～73%。这些差异反映了 DNA 物理性质的不同。

表 2.3　不同 DNA 的 G+C 相对含量

DNA 来源	百分数 (G+C) /%
盘基网柄菌 *Dictyostelium*（黏液菌）	22
化脓性链球菌 *Streptococcus pyogenes*	34
疫苗病毒 Vaccina virus	36
蜡状芽孢杆菌 *Bacillus cereus*	37
巨大芽孢杆菌 *B. megaterium*	38
流感嗜血杆菌 *Haemophilus influenzae*	39
酿酒酵母菌 *Saccharomyces cerevisiae*	39
小牛胸腺 Calf thymus	40
大鼠肝 Rat liver	40
公牛精子 Bull sperm	41
肺炎链球菌 *Streptococcus pneumoniae*	42
小麦胚 Wheat germ	43
雏鸡肝 Chicken liver	43
小鼠脾 Mouse spleen	44
鲑鱼精子 Salmon sperm	44
枯草芽孢杆菌 *B. subtilis*	44
T1 噬菌体 T1 bacteriaphage	46
大肠杆菌 *Escherichia coli*	51
T7 噬菌体 T7 bacteriaphage	51
T3 噬菌体 T3 bacteriaphage	53
链孢酶 *Neurospora crassa*	54
绿脓杆菌 *Pseudomonas aeruginosa*	68
藤黄八叠球菌 *Sarcina lutea*	72
溶壁微球菌 *Micrococcus lysodeikticus*	72
疱疹病毒 Herpes simplex virus	72
草分枝杆菌 *Mycobacterium phlei*	73

Source: From Davidson, *The Biochemistry of the Nucleic Acids*, 8th ed. revised by Adams et al., Lippencott.

当 DNA 溶液被充分加热时，结合两条链的非共价力被削弱直至最后被打断。此时两条链在 **DNA 变性**（DNA denaturation）或 **DNA 熔解**（DNA melting）过程中分离。DNA 链一半变性时的温度称为**熔解温度**（melting temperature）或 T_m。图 2.17 是肺炎链球菌 DNA 的熔解曲线。链分离或熔解的量可通过 DNA 溶液在 260nm 处的光吸收来检测。碱基中的电子结构使核酸在这个波长吸收光，但是当两条 DNA 链结合时，碱基的紧密接近会猝灭一些吸收。当 DNA 链分离时，这种猝灭作用消失，吸光率升高 30%～40%，这种现象称为**增色转换（效应）**（hyperchromic shift）。曲线的急剧上升表明温度到达 T_m 之前，链结合得很紧，到达 T_m 之后迅速分离。

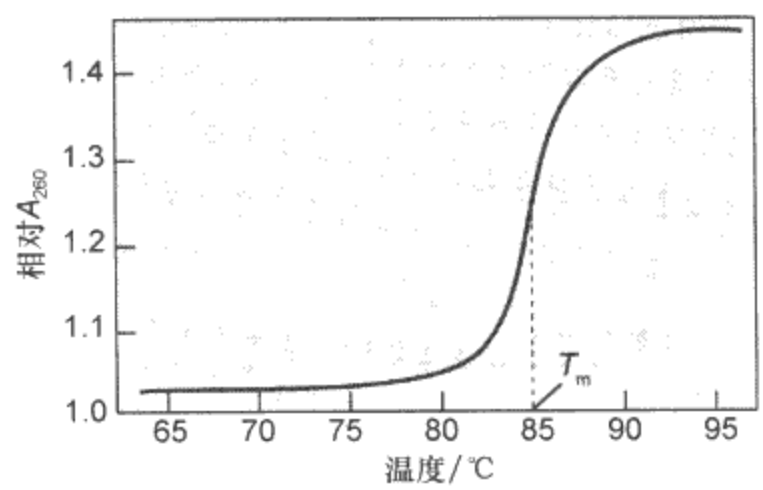


图 2.17　肺炎链球菌（*Streptococcus pneumoniae*）DNA 变性曲线。加热 DNA，其变性以 260nm 处吸光率的上升来检测。变性温度规定为完全熔解温度的一半或 T_m。这条 DNA 的 T_m 值约为 85℃。(Adapted from P. Doty, *The Harvey Lectures* 55: 121, 1961.)

DNA 链的 GC 含量对它的 T_m 值有显著影响。如图 2.18 所示，DNA 的 GC 含量越高，其 T_m 值就越高。为什么会这样？回忆使两条 DNA 链结合的力之一是氢键结合，再回想 G-C 有 3 个氢键，而 A-T 只有 2 个氢键。因此，富含 GC 的 DNA 双链比富含 AT 的 DNA 双链结合得更紧密。想想两对拥抱的蜈蚣，一对各有 200 条腿，另一对各有 300 条腿，显然后一对更难被分开。

加热并非 DNA 变性的唯一方法，二甲基亚砜、甲酰胺等有机溶剂或高 pH 都可破坏 DNA 链之间的氢键，促使其变性。降低 DNA 溶液中的盐浓度，移除可屏蔽两条链间负电荷的离子，也有助于变性。在很低的离子强度下，这些负电荷的互斥力很强，足以使 DNA 在相对低的温度下变性。

DNA 的 GC 含量同样影响它的密度。图 2.19 显示用 CsCl 溶液密度梯度离心时（见第 20 章）GC 含量与密度之间的直接线性关系。密度依赖碱基组

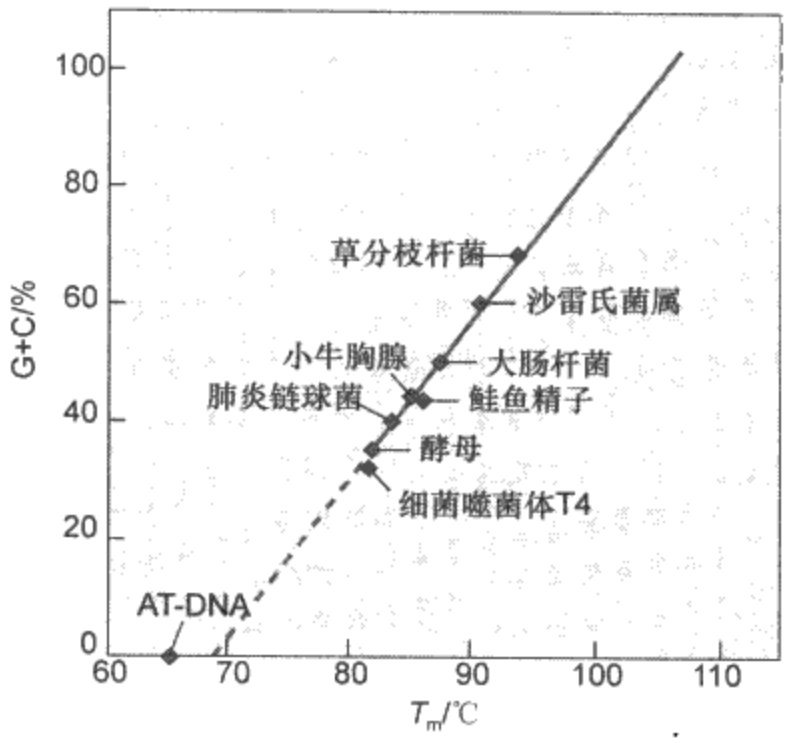


图 2.18　DNA 溶解温度与 GC 含量之间的关系。AT-DNA 是指人工合成的全部由 A 和 T 组成的 DNA（GC 含量 =0）。（Adapted from P. Doty, *The Harvey Lectures* 55：121，1961.）

成的部分原因是真实的，即 A-T 碱基对的摩尔体积较 G-C 碱基对要大，但部分原因可能是 CsCl 密度方法产生的假象：G-C 碱基对与 CsCl 结合的倾向性比 A-T 碱基对更强，使其观测密度看上去高于相对的实际密度。

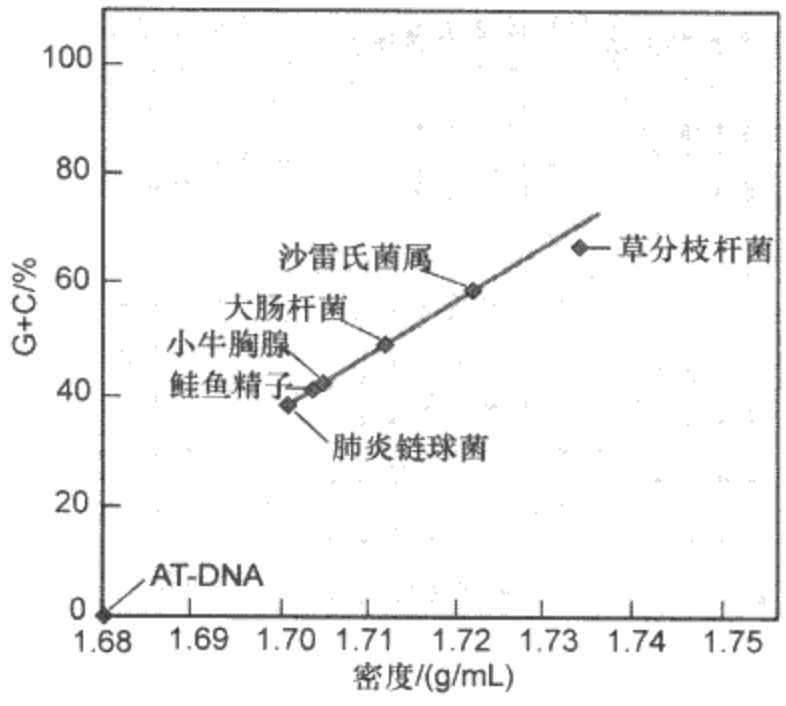


图 2.19　不同来源的 DNA 中 GC 含量与密度之间的关系。AT-DNA 是指人工合成的纯 A＋T 组成的 DNA；其 GC 含量为零。（Adapted from P. Doty, *The Harvey Lectures* 55：121，1961.）

小结　天然 DNA 中 GC 含量在少于 25% 和接近 75% 之间变化，这对 DNA 的物理性质影响强烈，特别是它的熔解温度和密度都随 GC 含量呈线性增加。DNA 的熔解温度（T_m）是一半双链解离或变性时的温度。低离子强度、高 pH 和有机溶剂都能促进 DNA 变性。

解离的 DNA 链的再结合　一旦 DNA 双链分离，在适当条件下它们可再次结合，这一过程称为**退火**（annealing）或复性（renaturation）。多种因素可影响复性效率。最重要的 3 种因素如下。

1. 温度　DNA 复性的最佳温度比其 T_m 值低约 25℃。这个温度低到不会促进变性，但也高到使 DNA 分子迅速扩散，削弱错配序列与短的链内碱基配对区域间的瞬时键合。表明变性后迅速冷却可防止复性。实际上，使变性 DNA 保持变性状态的常用方法是将热的 DNA 溶液迅速插入冰水中，这个方法叫做淬火（quenching）。

2. DNA 的浓度　溶液中 DNA 的浓度也很重要。在合理的范围内，浓度越高，两条互补链在给定时间内就越有可能相遇。换句话说，浓度越高，复性越快。

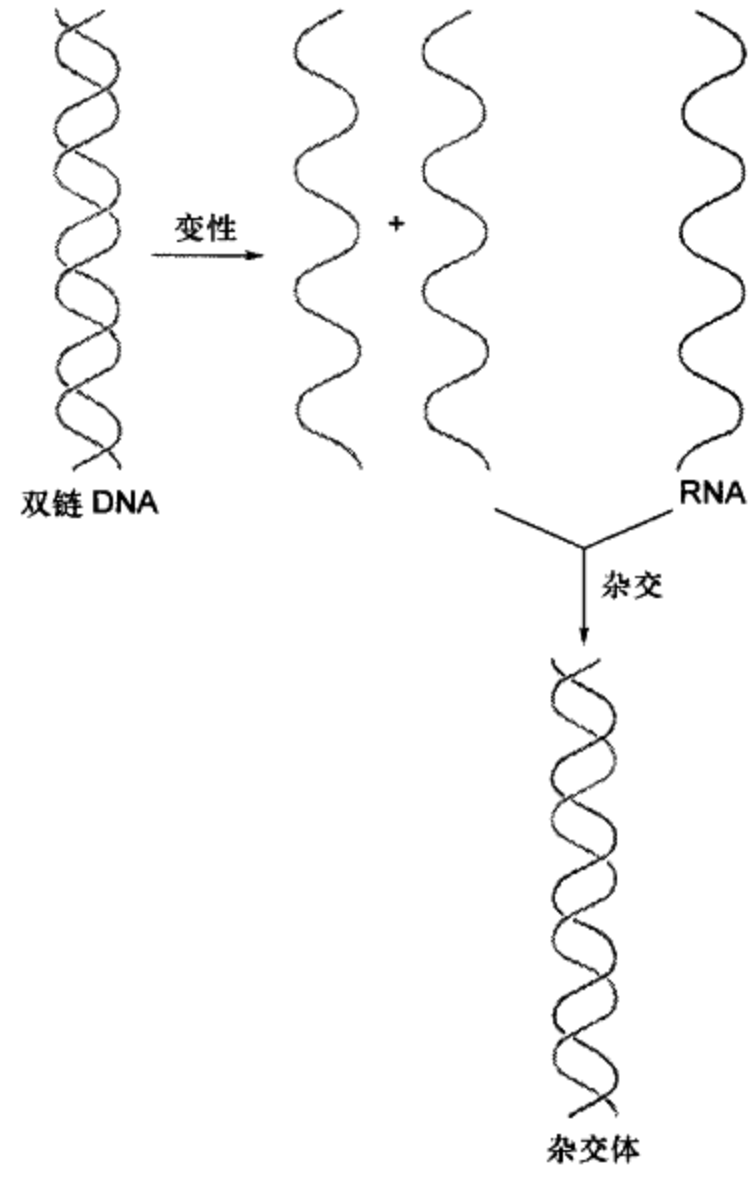


图 2.20　DNA 与 RNA 的杂交。首先将左上角的 DNA 变性，分开成两条 DNA 单链（蓝色）。然后，将其与 RNA 链混合，该 RNA 链只与其中一条 DNA 链互补。杂交反应在相对较高的温度下发生，有利于 DNA-RNA 杂交双链的形成。杂合链中有一条 DNA 链（蓝色）和一条 RNA 链（红色）。

3. 复性时间　显然，复性时间越长，复性越多。

小结　分离的 DNA 链可以进行复性或退火。多种因素影响复性，其中有温度、DNA 浓度和复性时间。

两种不同多聚核苷酸链的杂交　迄今为止，我们只是将两条分开的 DNA 链简单地再结合在一起，但是还有其他可能性。例如，一条 DNA 链与一条 RNA 链也可结合形成双螺旋。如果将基因的两条链分开，其中一条与其互补的 RNA 链放在一起，就会发生这种情况（图 2.20）。我们不再将其称为退火，而是**杂交**（hybridization），因为两种不同的核酸在一起形成杂合链（hybrid）。杂交的两条链不一定必须像 DNA 与 RNA 那样大的差异，两个不同的具有互补或几乎互补的 DNA 链放在一起，仍然称之为杂交——只要两条链来源不同。两个互补链的差别还可以非常小，如一条链可能有放射性而另一条没有。我们将在本书后面看到，杂交是极有价值的技术。实际上，杂交对分子生物学的重要性再怎么夸大都不为过。

不同大小和形状的 DNA

表 2.4 是几种生物的单倍体基因组和病毒的大小。有 3 种方法可表示基因组大小：分子质量、碱基数和碱基长度。当然，这些指标是相关的。我们已经知道怎样将碱基对数目转化为长度，每一圈螺旋（螺距）约有 10.4 个碱基对，长为 33.2Å。将碱基对数转化为分子质量，只需乘以 660（平均一个核苷酸对的分子质量）。

如何测量这些大小呢？对于短的 DNA 很容易。例如，PM2 噬菌体 DNA 含有一条双链环状 DNA。怎么知道它是环状的呢？最直接的方法就是看一下。用电子显微镜就能看到，但首先必须处理 DNA 使其能阻止电子并呈现在显微图像上，就像骨骼阻碍 X 射线从而在 X 射线照片上显现一样。最普通的方法是用铂等重金属包裹 DNA，将 DNA 放在一个电子显微镜网格上同时从很小的角度用小金属弹轰击，使金属堆积在 DNA 周围，就像栅栏后面的白雪。旋转网格上的 DNA 使其四周都变成阴影。现在金属将电子显微镜中的电子阻止并使 DNA 显现出来了，就像在黑暗背景下的光线。翻转图像产生如图 2.21 的照片，这是 PM2 DNA 两种形式的电子显微照片：开环的（左下）和**超螺旋**（supercoil）（右上），超螺旋的 DNA 自身盘绕很像一条扭曲的橡皮筋。利用这种图片就可测定 DNA 的长度了，如果在同一张图中含有已知长度的标准 DNA，则会使测定更为精确。DNA 的大小还可以用凝胶电泳估计，我们将在第 5 章介绍。

小结　天然 DNA 的大小在几千碱基到数亿碱基之间变化。小 DNA 的片段大小可直接用电子显微镜检测，这个技术也可揭示 DNA 是环状的、线性的还是超螺旋的。

表 2.4　各种大小的 DNA

来源	相对分子质量	碱基对/bp	长度
亚细胞遗传系统			
SV40（哺乳动物病毒）	3.5×10^6	5226	1.7μm
噬菌体 ΦX174（双链形式）	3.2×10^6	5386	1.8μm
噬菌体 λ	3.3×10^7	4.85×10^4	13μm
噬菌体 T2 或 T4	1.3×10^8	2×10^5	50μm
人类线粒体	9.5×10^6	16 596	5μm
细菌			
嗜血杆菌	1.2×10^9	1.83×10^6	620μm
大肠杆菌	3.2×10^9	4.64×10^6	1.6mm
沙门氏菌	8×10^9	1.1×10^7	3.8mm
真核生物（单倍体细胞核的含量）			
酵母	7.9×10^9	1.2×10^7	4.1mm
红色面包霉	约 1.9×10^{10}	约 2.7×10^7	约 9.2mm
黑腹果蝇	约 1.2×10^{11}	约 1.8×10^8	约 6.0cm
小鼠	约 1.5×10^{12}	约 2.2×10^9	约 750cm
人类	约 2.3×10^{12}	约 3.2×10^9	约 1.1m
玉米	约 4.4×10^{12}	约 6.6×10^9	约 2.2m
蛙	约 1.4×10^{13}	约 2.3×10^{10}	约 7.7m
百合	约 2×10^{14}	约 3×10^{11}	约 100m

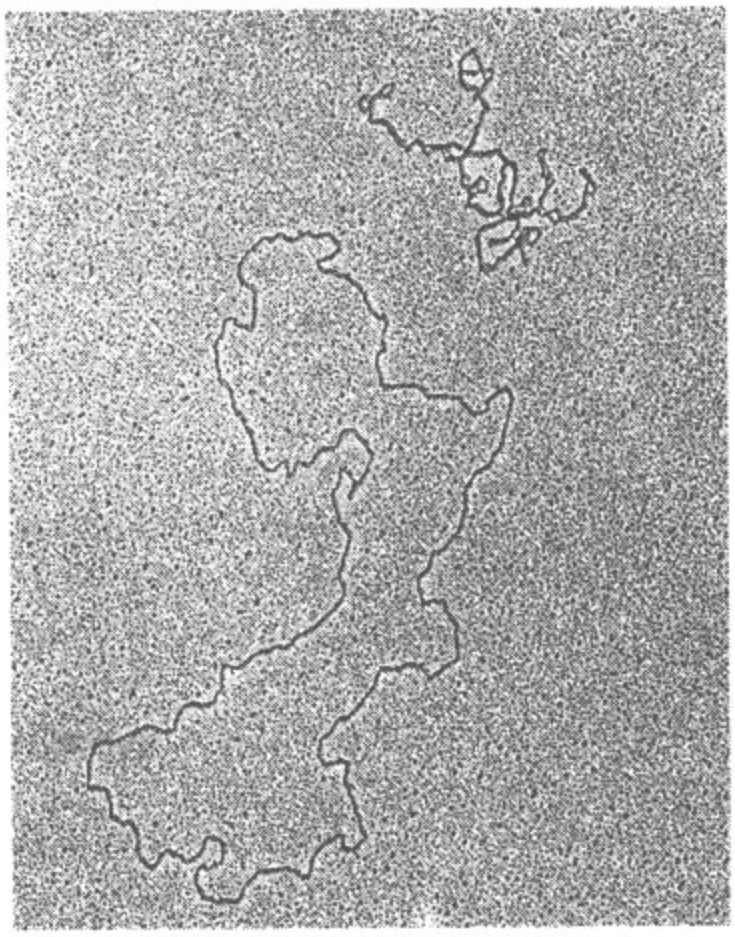


图 2.21　噬菌体 PM2 DNA 的电子显微照片。左下方是开环形式，右上方是超螺旋。（*Source*：© Jack Griffith.）

问题解答

问题 3

噬菌体 P1 是一条含有 91 500bp（91.5kb）的双链 DNA。（a）这条 DNA 含有多少个完整的双螺旋转数？（b）该 DNA 有多少微米（$1\mu m = 10^4 Å$）？（c）其分子质量是多少？（d）含有多少磷原子？

解答

（a）已知双螺旋每转一圈含有 10.4bp 碱基。因此，我们所要做的就是用碱基总数除以 10.4

91 500bp×1 个螺旋转/10.4bp≈ 8798 个螺旋转

（b）沿螺旋轴，碱基对之间的距离约 3.32Å，或者 $3.32\times10^{-4}\mu m$，所以只需把碱基对总数乘以 3.32×10^{-4} 即可

91 500bp×$3.32\times10^{-4}\mu m$ /bp ≈30.4μm

（c）一个碱基对的分子质量约为 660 道尔顿（Da），所以用碱基对总数乘以 660Da

91 500bp×660Da/bp=6.0×10^7 Da（6.0×10^{-4} kDa）

（d）每个碱基对有两个磷原子，每条链有一个。因此，只要用碱基对总数乘以 2 即可

91 500bp×2 磷原子/bp=183 000 个磷原子

DNA 大小与遗传容量之间的关系　一个给定的 DNA 中有多少个基因？仅从 DNA 大小是不可能回答的，因为我们不知道一条给定的 DNA 中有多少序列用于基因或基因间隔、基因内有多少个插入序列。但是可以估计一条 DNA 能够容纳基因的上限数目。先假定要讨论的基因是编码蛋白质的，在第 3 章和其他章将看到许多基因只编码 RNA，在这里不考虑它们。还要假定一个普通蛋白质的分子质量约为 40 000Da，这代表多少个氨基酸呢？氨基酸的分子质量不同，但是平均约为 110Da。为简化计算，假设平均为 110，意味着普通蛋白质含有 40 000/110，约 364 个氨基酸。因为每个氨基酸需要 3bp 的 DNA 编码，含有 364 个氨基酸的蛋白质需要一条长约 1092bp 的基因。

从表 2.4 列出的 DNA 中看出，大肠杆菌染色体 4.2×10^6 bp，可编码约 4200 个普通蛋白质。感染大肠杆菌的 λ 噬菌体只有 4.85×10^4 bp，只能编码 44 个蛋白质。表 2.4 中最小的双链 DNA 是 ΦX174，只有 5375bp，原则上只够编码约 5 个蛋白质，但是噬菌体可通过重叠基因挤出一些额外信息。

问题解答

问题 4

表 2.4 显示噬菌体 T2 和 T4 DNA 含有 2×10^5 bp。这些噬菌体能含有多少个平均大小的基因（编码蛋白的相对分子质量约为 40 000）？

解答

因为一个氨基酸的平均分子质量是 110Da，分子质量为 40 000Da 的蛋白质约含有 364 个氨基酸。3 个碱基编码一个氨基酸，所以每个蛋白质将由 1092 个碱基编码。因此，在一条含有 2×10^5 bp 的 DNA 中，中等大小的基因数是 $2\times10^5/1.09\times10^3 = 1.83\times10^2$ 或 183。

DNA 的含量与 C 值悖论　你可能会预测像脊椎动物那样复杂的生物体要比酵母那样简单的生物体需要更多的基因，因此它们应当有更高的 **C 值**（C-value），即每个单倍体细胞的 DNA 含量。大体上，你的预测是正确的。小鼠和人类单倍体细胞的 DNA 量是酵母单倍体细胞的 100 多倍。而酵母细胞的 DNA 比更简单的大肠杆菌多 5 倍。然而，生物体的物理复杂性与其细胞中 DNA 含量间的对应关系并不完美。以青蛙为例，本能地你不会认为两栖动物的 C 值比人类高，然而，青蛙每个细胞的 DNA 含量是人类的 7 倍，甚至更为夸张的是百合花每个细胞所含的 DNA 比人类细胞的多 100 倍！

这种令人困惑的情况称为 **C 值悖论**（C-value paradox）。考虑同一类群内的生物时，情况就变得更加难以解释。例如，有些两栖动物的 C 值比其他两栖动物高 100 倍，而显花植物的 C 值变化更大。这是否意味着某种高等植物的基因要比另一种高等植物多出 100 倍？这简直难以置信。问题是那些额

外的基因有什么用？为什么我们看不到这些生物在物理复杂性上的巨大差异。对C值悖论较可信的解释是具有很高C值的生物体只是具有大量多余的、非编码的DNA。这些额外DNA的功能，如果有的话，仍然是神秘的。

实际上，甚至哺乳动物的DNA也远远多于基因的需要。将我们的简单规则（用碱基对数除以1090）用于人类基因组，可得到最大基因数约为300万的估计，这个数目太高了。

已完成的人类基因组草图表明人类只含20 000～25 000个基因，这意味着人类细胞含有的DNA要比看上去所需要的多100倍。在真核生物基因中，多数额外DNA是在间插序列中发现的（第14章），其余的则在基因之外的非编码区。

小结 在细胞或病毒中，DNA含量与基因数目间有一定的相关性。然而，这种相关性在几个亲缘关系相近的生物中被打破了，其单倍体细胞的DNA含量（C值）变化很大。C值悖论在某些生物或许可以通过额外非编码DNA而不是额外基因来解释。

总结

所有真核生物的基因都是由DNA构成的，一些病毒的基因由RNA构成。DNA和RNA是由称之为核苷酸的基本单位组成的链状分子。DNA具有双链螺旋结构，其糖-磷酸骨架在外侧，碱基对在里面。碱基以特殊的方式配对：腺嘌呤（A）与胸腺嘧啶（T）配对，鸟嘌呤（G）与胞嘧啶（C）配对。当DNA复制时，母链分开，并分别作为模板合成一条新的互补链。

天然DNA的G＋C含量在22%～73%变化，对DNA的物理性质，特别是其熔解温度产生强烈影响。DNA的熔解温度（T_m）是一半双链解离或变性时的温度。分开的DNA链可被复性或退火。在称之为杂交的过程中，不同来源的多聚核苷酸互补链（RNA或DNA）能形成双螺旋。天然DNA长度变化很大。小DNA的片段大小能通过电子显微镜测定。

在细胞或病毒中DNA含量与基因数目间有一定的相关性。但是，在几个亲缘关系相近的生物中这种相关性不成立，这些生物每个单倍体细胞的DNA含量（C值）变化很大。C值悖论或许可以通过某些生物的额外非编码DNA来解释。

复习题

1. 比较和对比Avery及同事以及Hershey和Chase为证明DNA是遗传物质所采用的实验方法。

2. 图示脱氧核苷单磷酸的一般结构，详细显示糖结构，标出碱基和磷酸连接的位置，标出脱氧的位置。

3. 图示连接两个核苷酸的磷酸二酯键的结构。清楚显示磷酸二酯键中两个糖的位点。

4. 哪个DNA嘌呤与另一条DNA链上的配对碱基形成3个氢键？哪个形成2个氢键？哪种DNA嘧啶与其配对碱基形成3个氢键？哪个形成2个氢键？

5. 图示典型的DNA熔解曲线图，标出坐标轴并指出熔解温度。

6. 图示DNA的GC含量与其熔解温度间的关系，对这种关系的解释是什么？

7. 图示DNA的GC含量与其密度间的关系，对这种关系的解释是什么？

8. 图示说明核酸分子杂交的原理。

分析题

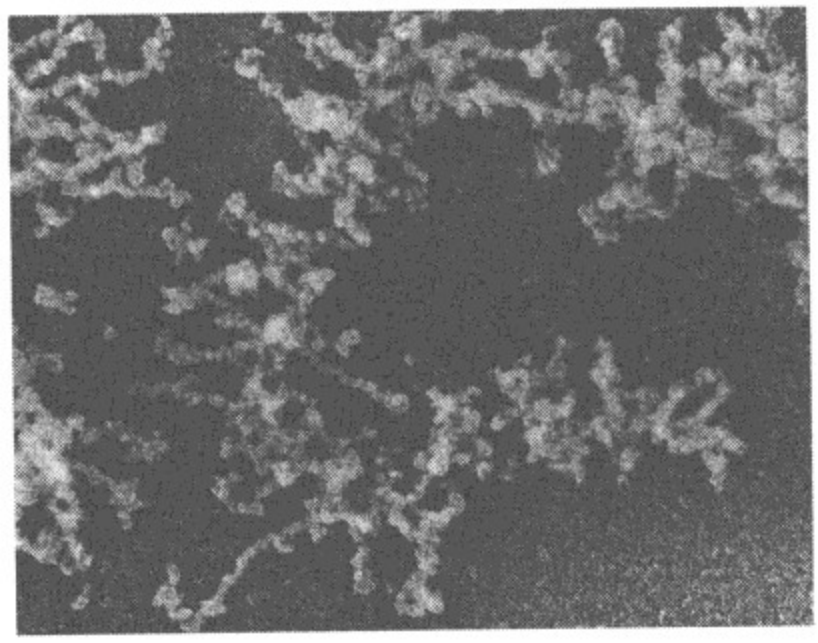
1. 人类单纯疱疹I型病毒基因组双链DNA的分子质量约为1.0×10^5kDa。

（a）这个病毒有多少个碱基对？（b）有多少个完整的双螺旋？（c）DNA的长度是多少微米？

2. 一个含有12 000bp的DNA基因组能编码多少个平均大小的蛋白质，假设基因不重叠？

翻译 马 纪 校译 张富春

第3章　基因功能简介



色彩增强的 DNA 链扫描电镜照片，显示转录细节（×61 500）。

© *Biophoto Associates/Photo Researchers, Inc.*

正如我们在第 1 章所学过的，基因参与以下 3 种主要活动。

1. 基因是个信息库，携带有合成生命关键分子——RNA 的信息。RNA 的碱基序列直接来自基因的碱基序列，而大多数 RNA 又作为模板合成另一种关键的细胞分子——蛋白质。以 DNA 为蓝图产生 RNA 和蛋白质的过程叫做基因表达。第 6～19 章将讨论基因表达的各个方面。

2. 基因能被复制。这种复制是非常准确保守的，使遗传信息能一代一代传下去。第 20 章、第 21 章将讨论基因的复制。

3. 基因可接受偶然的改变或突变，使生物体得以进化。有时这些变化包括重组，即染色体之间或一条染色体上位点之间 DNA 的交换。重组还包括基因组中 DNA 片段（转座元件）（transposable element）从一个位置转移到另一个位置。第 22 章、第 23 章将讨论重组和转座因子。

本章简单介绍了基因的 3 种特性，在后面的章节将提供一些背景知识以利于深入理解这些特性。

3.1　储存信息

我们先对基因表达的过程做简要的概述，然后介绍蛋白质结构，概括基因表达的两个步骤。

基因表达的总体过程

从 DNA 基因的信息到蛋白质的合成是一个两步过程。第一步是 RNA 的合成，这条 RNA 与 DNA 的一条链互补，称为**转录**（transcription）。第二步称为**翻译**（translation），RNA 的信息用于合成多肽。这种信息性的 RNA 称为**信使 RNA**（mRNA），表示它像信使一样携带信息从基因到细胞的蛋白质工厂。

与 DNA 和 RNA 一样，蛋白质也是多聚物——长的链状分子。蛋白质链中的单体称为氨基酸。DNA 与蛋白质的信息关系为：DNA 基因的 3 个核苷酸代表蛋白质的一个氨基酸。

图 3.1 总结了蛋白质编码基因的表达过程及关于 DNA 链的术语。mRNA 与上方（蓝色）的 DNA 链有相同的序列（除了 U 代替 T）。mRNA 含有生成多肽的信息，所以我们说它“编码”（code for 或 encode）多肽。这里 mRNA 编码甲硫氨酸-丝氨酸-天冬酰胺-丙氨酸，缩写为 Met-Ser-Asn-Ala。在这条 mRNA 上甲硫氨酸密码子为 AUG，丝氨酸、天冬酰胺和丙氨酸的密码子分别为 AGU、AAC 和 GCG。

由于下方的 DNA 链与 mRNA 互补，以其为模板生成 mRNA，因此称这条链为**模板链**（template strand）或转录链，而上方这条链则为**非模板链**（nontemplate strand）或非转录链。由于非模板链与相应的 mRNA 在本质上有相同的编码特征，许多遗传学家又称它为编码链（coding strand），与它相对的链为非编码链（anticoding strand）。同样，由于上方链与 mRNA 意义相同，也称为正义链（sense

strand)，下方的链称为反义链（antisense strand）。但是遗传学家使用的“编码链”和“正义链”却是完全相反的情况。为避免混乱，本书用模板链和非模板链两个术语。

蛋白质结构

为了理解基因的表达，并且大多数基因的终产物是蛋白质，我们先简略看一下蛋白质的性质。像核酸一样，蛋白质是由亚单位组成的链状多聚物。DNA 和 RNA 的链环是核苷酸，而蛋白质的链环是氨基酸（amino acid）。DNA 只有 4 种核苷酸，蛋白质却有 20 种不同的氨基酸，其结构如图 3.2 所示。每个氨基酸有一个氨基（$-NH_3^+$）、一个羧基（$-COO^-$）、一个氢原子（H）和一个侧链。两个氨基酸的唯一差异只是它们的侧链不同。因此，蛋白质

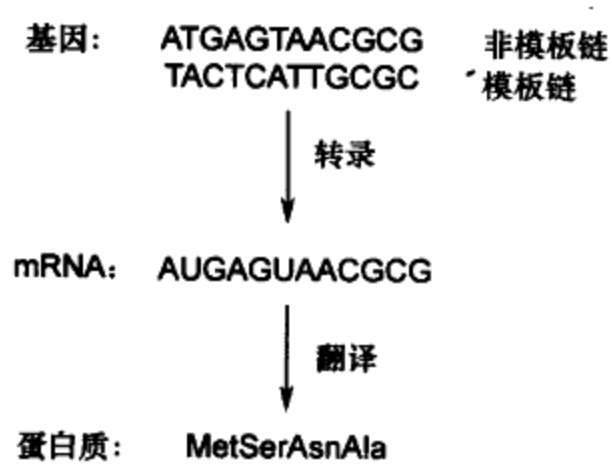


图 3.1 基因表达概要。第一步，转录：模板链转录成 mRNA。非模板链与 mRNA 有相同的序列（除 T-U 变化之外）。第二步，mRNA 翻译成蛋白质。这个小“基因”只有 12 个碱基对，编码 4 个氨基酸（四肽）。真实的基因大得多。

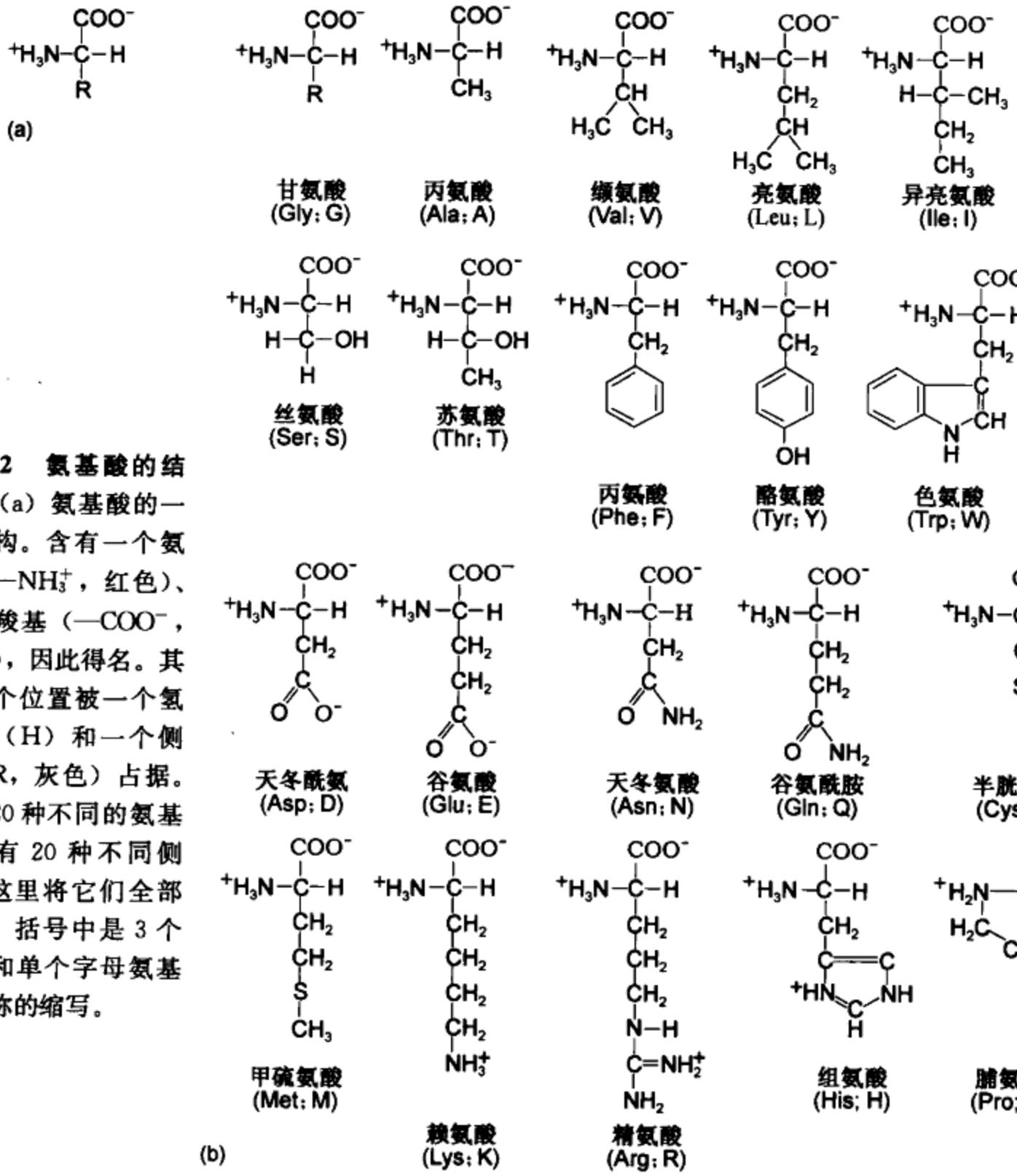


图 3.2 氨基酸的结构。（a）氨基酸的一般结构。含有一个氨基（$-NH_3^+$，红色）、一个羧基（$-COO^-$，蓝色），因此得名。其他两个位置被一个氢原子（H）和一个侧链（R，灰色）占据。（b）20 种不同的氨基酸就有 20 种不同侧链，这里将它们全部列出。括号中是 3 个字母和单个字母氨基酸名称的缩写。

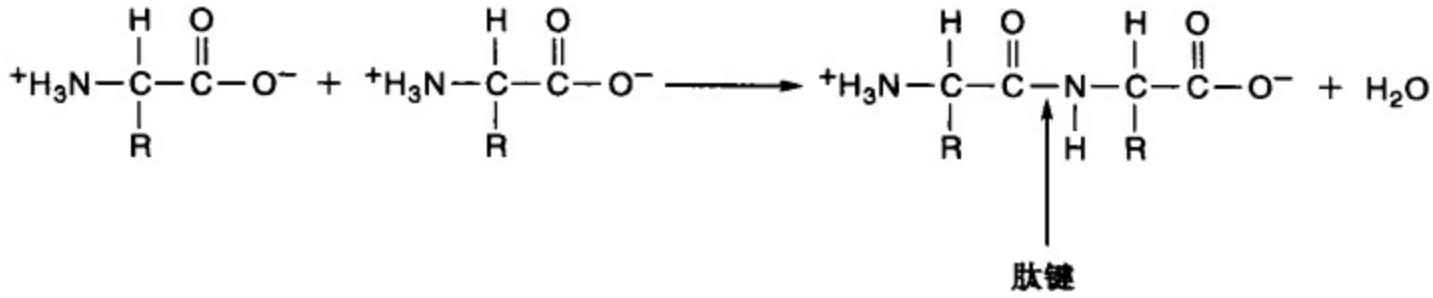


图 3.3　肽键的形成。有侧链 R 和 R′的两个氨基酸通过第一个氨基酸的羧基与第二个氨基酸的氨基形成 1 个二肽，2 个氨基酸通过肽键连接，并产生了 1 分子水。（译者注：图中第二个氨基酸 R 应为 R′）

正是由于氨基酸不同侧链的排列而赋予其独一无二的特性。蛋白质中氨基酸通过**肽键**（peptide bond）连接在一起（图 3.3），因此氨基酸链又被称为**多肽**（polypeptide）。一个蛋白质可由一个或多个多肽组成。

像 DNA 链一样，多肽链具有极性。图 3.3 右所示的二肽（dipeptide）（两个氨基酸连接在一起）的左端有一个自由的氨基，称为**氨基端**（amino terminus）或 **N 端**（N-terminus）；右端有一个自由的羧基，称为**羧基端**（carboxyl terminus）或 **C 端**（C-terminus）。

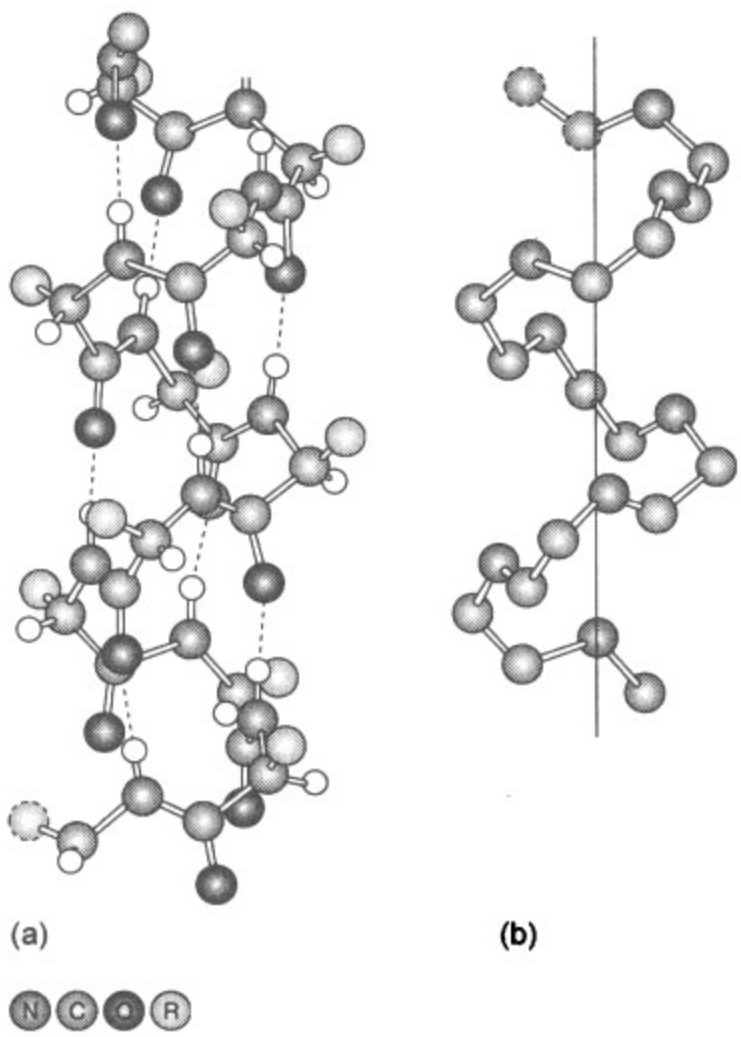


图 3.4　蛋白质二级结构示例：α 螺旋。（a）显示氨基酸在螺旋中的位置，螺旋骨架用黑色和蓝色表示。虚线表示氨基酸氢原子和氧原子间的氢键。小白球表示氢原子。（b）α 螺旋的简化表示，只显示螺旋骨架中的原子。

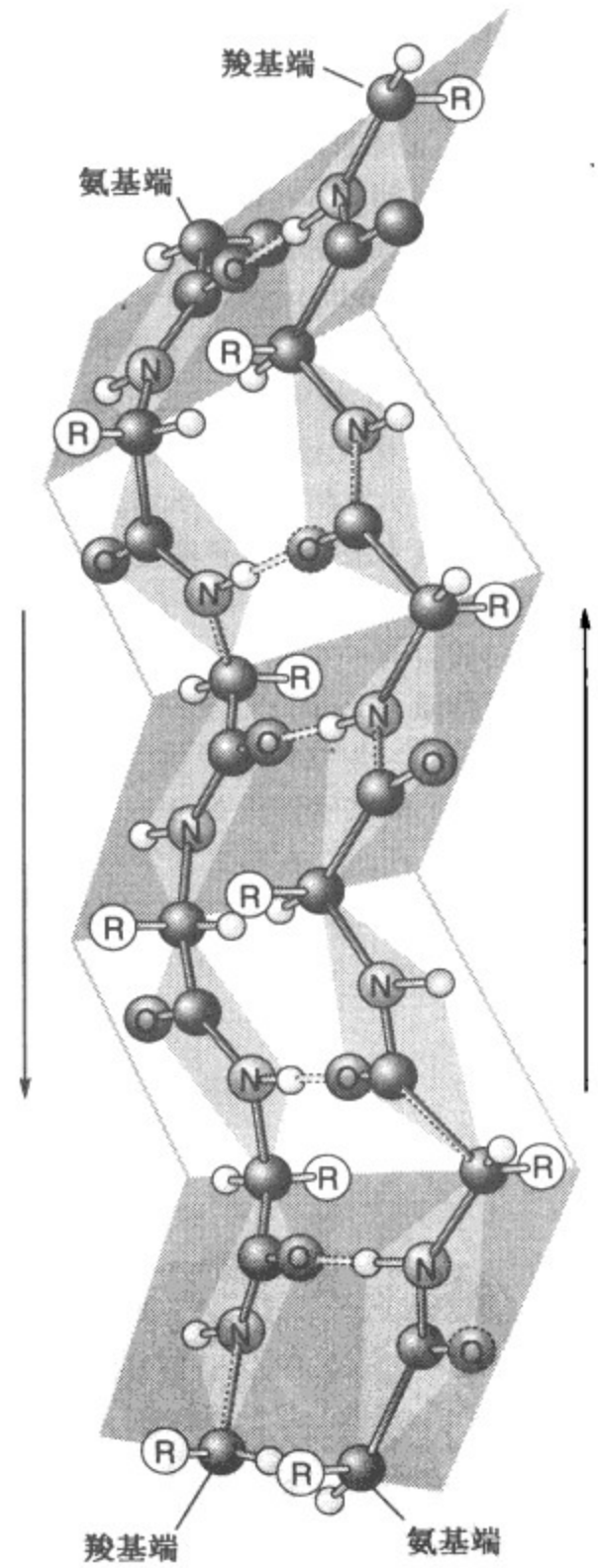


图 3.5　反向平行的 β 折叠。两条多肽链通过其间的氢键（虚线）并肩排列。绿色和白色平面显示 β 片是折叠的。两条链反向平行，一条链的 N 端与另一条链的 C 端在上方。箭头表示两条 β 链以相反方向从 N 端到 C 端。平行 β 折叠也存在，其 β 链按相同方向延伸。

氨基酸的线性顺序组成了蛋白质的**初级结构**（primary structure）。氨基酸之间相互作用的方式形成了蛋白质的**二级结构**（secondary structure）。α螺旋（α-helix）是一种常见的二级结构，它由相邻氨基酸间的氢键结合而产生（图 3.4）。另一种常见二级结构是β折叠（β-pleated sheet）（图 3.5），伸展的蛋白质链通过氢键作用一段接一段形成片状外观。丝蛋白富含β折叠结构。第三种二级结构只是一个转角，起连接α螺旋和β折叠作用。

多肽的三维空间形状称为**三级结构**（tertiary structure）。图 3.6 显示了肌红蛋白（myoglobin）是如何折叠成三级结构的。它含有二级结构的组成元件，特别是几个α螺旋。注意其大致的球形结构，多数多肽都具有这种形状，称为球蛋白（globular）。

图 3.7 是蛋白质结构的丝带模型。这是醋酸胍甲基转移酶（guanidinoacetate methyltransferase，GAMT）的三维结构。可以清楚地看到 3 种二级结构：α螺旋，用螺旋形丝带表示；β折叠，用并排的扁平箭头表示；结构元件之间的转角用细绳表示。图 3.7 中的球棍表示有两个小分子结合在蛋白质上。可用立体镜或“魔术眼”技术观察这个三维图。

肌红蛋白和 GAMT 都是单一的、大致的球状结构，有些蛋白质含有一个以上的紧密结构区域，这种区域叫做**结构域**（domain）。抗体（白细胞抵御感染时产生的蛋白质）为结构域提供了一个很好的例子。IgG 型抗体 4 个多肽链中的任一条都含有球状结构域，如图 3.8 所示。在第 9 章学习蛋白质-DNA 结合时，我们将看到结构域可含有共同的结构-功能**基序**（motif）。例如， 叫做锌指（zinc finger）的指

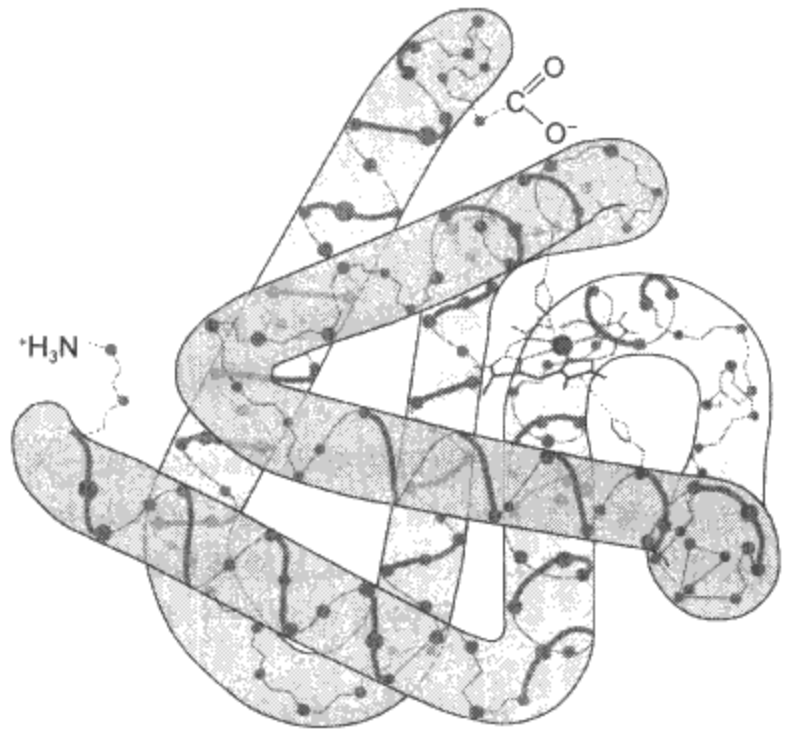


图 3.6　肌红蛋白的三级结构。α螺旋用绿松色螺锥表示。整个分子像一根香肠，盘旋成大致的椭圆或球形。血红素基团是红色的，结合蛋白质中的两个组氨酸（绿松色多边形）。

状基序参与结合 DNA。图 3.8 也描述了蛋白质的最高结构——**四级结构**（quaternary structure）。四级结构是复杂蛋白质中两条以上多肽链组合的方式。长期以来，人们设想蛋白质的氨基酸顺序决定了它的所有高级结构，就像本书中文字的线性顺序决定了文字、句子和段落结构。但这种类比过于简单了，大部分蛋白质离开其正常生活的细胞环境就不能正确折叠，这时除了蛋白质之外还需要一些细胞因子，在多肽合成期也必须经常发生折叠。

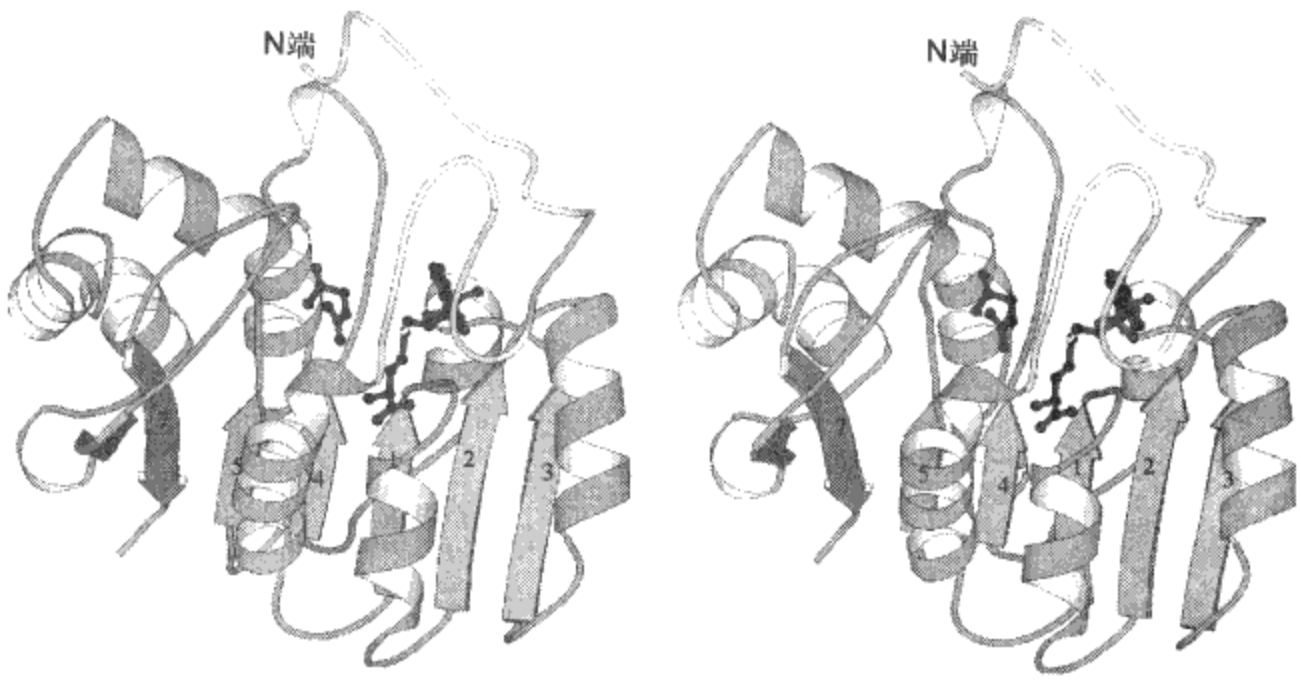


图 3.7　醋酸胍甲基转移酶（GAMT）的三级结构。二级结构元素包括α螺旋（螺旋形丝带）、β折叠（编号的平面箭头）、转角（细绳），显而易见。两个结合的分子（球棍图）是醋酸胍（左）和 S-腺苷甲硫氨酸（右），前者是酶的一个底物，后者是产物抑制剂。（*Source*: Reprinted with permission from Fusao Takusagawa, University of Kansas.）

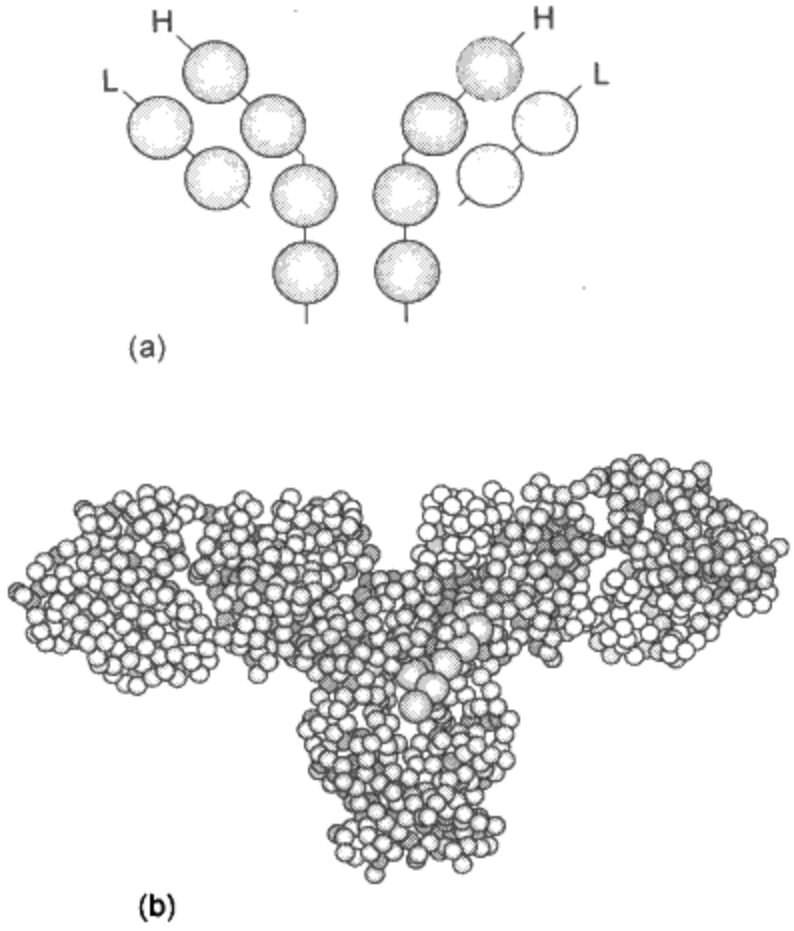


图 3.8　免疫球蛋白的球状结构域。（a）组成免疫球蛋白的 4 条多肽示意图：两条轻链（L）和两条重链（H）。轻链各含有 2 个球形结构域，重链各含有 4 个球形结构域。（b）免疫球蛋白的空间填充模型。颜色与（a）图相对应。黏附在蛋白质上的一个复杂糖分子用灰色表示。注意每个多肽的球状结构域，同时请注意 4 条多肽如何结合形成蛋白质的四级结构。

是什么作用力维持了蛋白质的正确形态？有些是共价键，但大部分是非共价键。多肽内和多肽间的主要共价键是半胱氨酸之间的二硫键（S—S），非共价键主要是疏水键和氢键。疏水性氨基酸集结在多肽的内部，或多肽间的结合面，以便避免与水接触［疏水（hydrophobic）即怕水的意思］。疏水作用在蛋白质的三级和四级结构中起主要作用。

小结　蛋白质是通过肽键连接起来的氨基酸多聚体。在多肽中氨基酸序列（初级结构）决定分子局部形态（二级结构）、整体形态（三级结构），以及与其他多肽间的相互作用（四级结构）。

蛋白质的功能

为什么蛋白质如此重要？因为有些蛋白质帮助细胞维持形状及完整性，有些作为激素（hormone）介导细胞间的信号转导。例如，胰腺分泌的胰岛素是肝细胞和肌肉细胞从血液中吸收葡萄糖的信号。蛋白质同样也能结合和运输物质，血红蛋白（hemoglobin）将氧从肺部传送到身体外周，肌红蛋白将氧储存在肌肉组织中。在本书中我们将多次看到，蛋白质也能控制基因的活动。蛋白质作为酶催化数百种生命所必需的化学反应。

基因与蛋白质之间的关系　关于基因-蛋白质关系的知识要回溯到 1902 年，当时一位名叫 Archibald Garrod 的医生注意到人类尿黑酸症（alcaptonuria）好像是由单个隐性基因引起的。幸运的是孟德尔的工作在 1900 年已被重新发现并为 Garrod 的观察提供了理论背景。尿黑酸症患者分泌过多的尿黑酸，使尿液颜色呈黑色。Garrod 分析，这种病是由于一种缺陷的代谢途径产生的化合物不正常积累所导致。在代谢途径的某个地方发生阻塞，导致中间代谢产物尿黑酸不正常积累到很高水平，就像一座大坝使水积聚一样。几年后，Garrod 推测这个问题来自于苯丙氨酸降解途径的缺陷（图 3.9）。

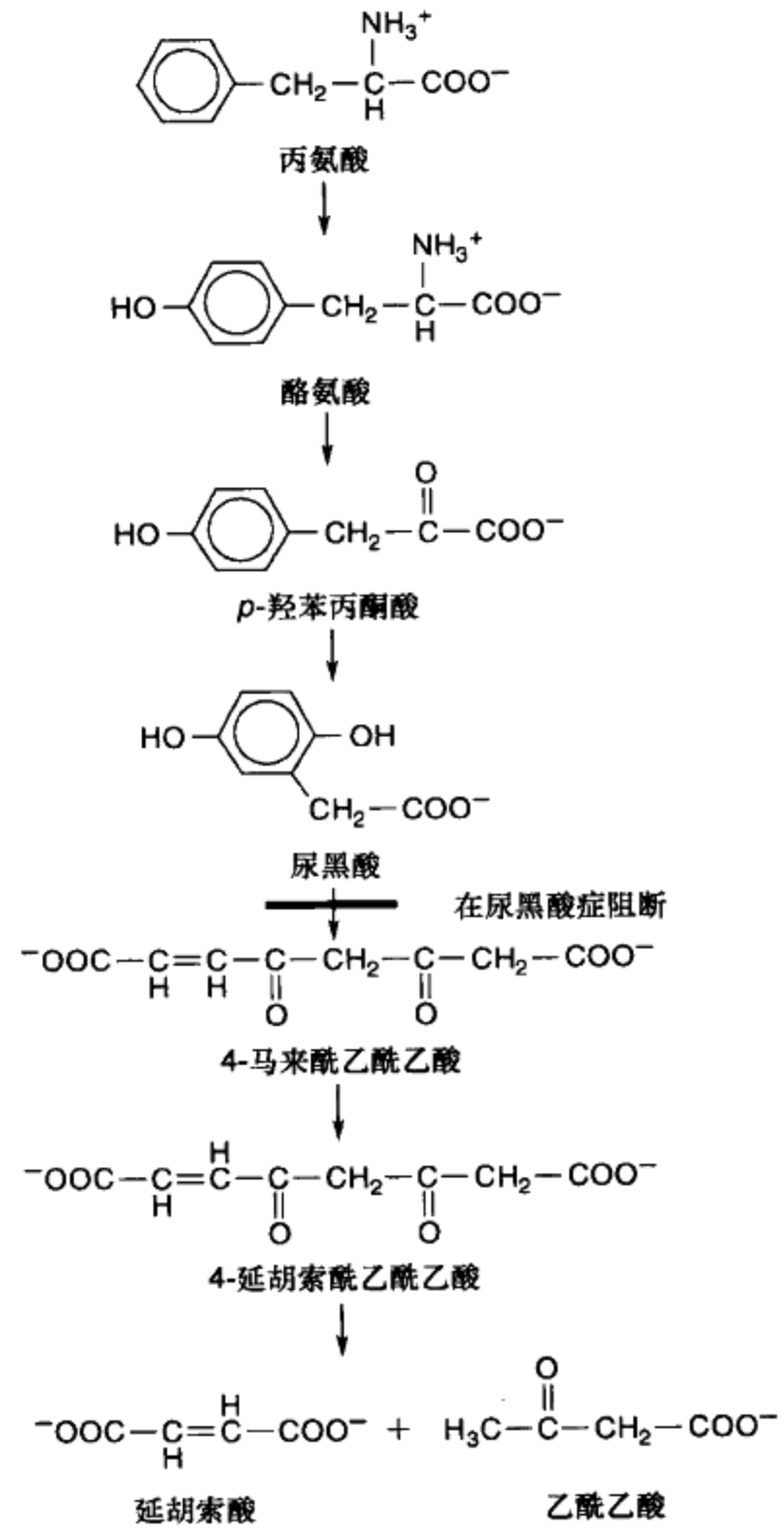


图 3.9　苯基丙氨酸降解途径。尿黑酸症患者缺乏将尿黑酸转化成 4-顺丁烯二酸单酰乙酰乙酸的酶。

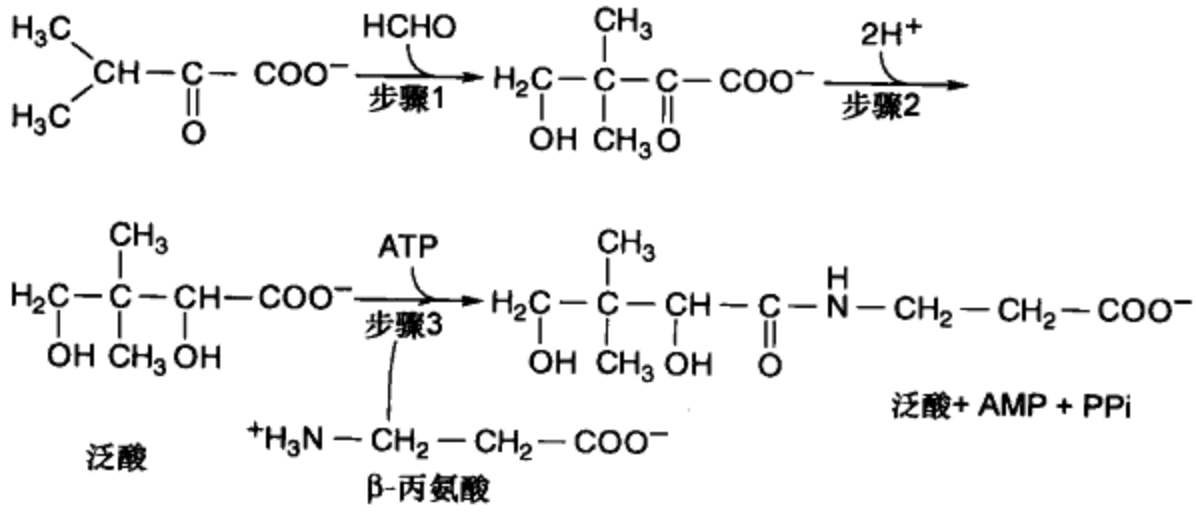


图 3.10　泛酸合成途径。最后一步（第 3 步），在 Beadle 和 Tatum 的一种突变体中由泛酸和 β-丙氨酸形成泛酸这一步被阻断了。催化该步骤的酶一定是缺陷性的。

当时代谢途径已被研究了多年并且知道是由酶控制的（酶催化），一个酶催化一个步骤。因此尿黑酸患者可能携带了一个缺陷型的酶。又因为该病以简单的孟德尔规律遗传，Garrod 推断某个基因可能控制该酶的产生。如果这个基因发生缺陷，就会产生一个缺陷的酶。这提示基因和蛋白质之间存在很重要的联系。

在 20 世纪 40 年代 George Beadle 和 Tatum 用普通面包**霉脉胞菌**（*Neurospora crassa*）做了进一步研究。他们先用 X 射线或放射线照射脉胞菌的子囊（形成孢子的部位）引起突变，然后从照射过的霉菌收集孢子并将其分别单独培养获得纯菌株。他们筛选了数千株才发现了一些突变体。突变株不能在只含有糖、盐、无机氮和生物素的基本培养基上生长，而野生型脉胞菌则很容易生长，突变株生长需添加一些额外的物质，如维生素。

接下来，Beadle 和 Tatum 对突变株进行了生化和遗传分析。通过每次将不同成分添加到突变体培养物中，他们追踪到了这个生化缺陷。例如，泛酸维生素合成的最后一步是将泛酸（pantoate）和 β-丙氨酸结合在一起（图 3.10）。“泛酸盐缺失”突变体能在含泛酸盐的培养基中生长，而不能在只含泛酸或 β-丙氨酸的培养基中生长。这证明在生化反应途径的最后一步（步骤 3）泛酸盐受阻，所以催化该步骤的酶一定是缺陷性的。

以上遗传分析只是浅易的。实际上脉胞菌是一种子囊菌（*ascomycete*），其两种不同配子型的细胞核融合并经减数分裂产生 8 个单倍体囊孢子存在于子实体中，称为子囊（ascus）。因此，突变株可与相反交配型（mating type）的野生株杂交产生 8 个孢子（图 3.11）。如果突变体表型是由单个基因突变导致的，那么其中 4 个孢子应该是突变体，另外 4 个是野生型。收集孢子并分别培养，检测所得霉菌的表型。果然，他们发现 8 个孢子中有 4 个产生了突变霉菌，证明突变表型是由单个基因控制的。这种情况可反复发生，由此研究人员得出结论：生化反应途径中的每个酶都是由一个基因控制的。

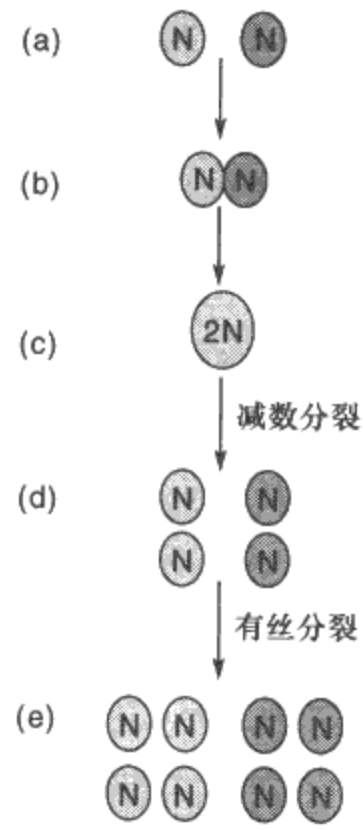


图 3.11　脉胞菌孢子的形成。（a）两个单倍体核子，野生型（黄色）和突变体（蓝色）在霉菌未成熟的子实体中汇聚。（b）两个核子开始融合。（c）融合完成，双倍体核子（绿色）形成。一套单倍染色体来自野生型，一套来自突变体。（d）减数分裂发生，产生 4 个单倍体核子。如果突变体的表型由单基因控制，则这些孢子中会有 2 个（蓝色）含突变体等位基因，2 个（黄色）含野生型等位基因。（e）最后，有丝分裂发生，产生 8 个单倍体核子，每个核子都将成为一个囊孢子。其中 4 个（蓝色）含突变等位基因，4 个（黄色）含野生型等位基因。如果突变的表型由多基因控制，结果将更复杂。

随后的研究表明许多酶含有一个以上多肽链，并且每个多肽是由一个基因编码的。这就是**一个基因/一个多肽**的假说。如第 1 章所提到的，这个假说还需要修正以便解释编码 RNA 的基因，如 tRNA 和 rRNA 基因。数十年来，人们认为这种 RNA 基因的数量很小，不超过 100 个。但是 21 世纪已经目

睹了对非编码 RNA 发现的爆炸式增长，目前仅在人类中就发现了数千种。其中一些 RNA 也许没有任何功能，不满足对基因产物的定义，但是也有许多具有可证明的重要功能。因此，“基因”这个词的定义已经变得更复杂、更具有争议了。我们现在认识了重叠基因（overlapping gene）、基因内基因（gene within gene）、片段基因（fragment gene）以及可能更异乎寻常的基因。在本书后面将讨论这些复杂性，在本章的剩余部分，我们将讨论“传统”基因，即编码蛋白质基因的表达。

小结 大多数基因含有合成一条多肽链的信息。

信使 RNA 的发现

在沃森和克里克的 DNA 模型发表后的几年里，mRNA 将遗传信息从基因带到核糖体的概念分为几个阶段建立起来了。1958 年，克里克本人提出 RNA 是遗传信息的中间载体。他的假设部分是基于 DNA 位于真核细胞的细胞核，而蛋白质在细胞质中合成这样一个事实，这就意味着有某种东西携带信息从一处转到另一处。克里克注意到核糖体含有 RNA，暗示这种核糖体 RNA（rRNA）是信息的载体。但是 rRNA 是核糖体的组成部分，不可能离开核糖体。因此，克里克的假设暗示每个核糖体以其自己的 rRNA 重复生产同一种蛋白质。

François Jacob 及同事提出了另一种假设，认为是非特异性核糖体翻译一种叫做**信使**（messenger）的不稳定 RNA。他们认为信使是将遗传信息从基因带到核糖体的独立 RNA。1961 年，Jacob 与 Sydney Brenner 和 Matthew Meselson 一起发表了关于信使假说的证据。研究采用的噬菌体（T2）与 Hershey 和 Chase 10 年前揭示基因由 DNA 组成所用的噬菌体一样。实验的前提是：噬菌体 T2 感染大肠杆菌，使宿主由合成细菌蛋白转而合成噬菌体蛋白。如果克里克的假设正确，那么对噬菌体蛋白合成的转换应伴随装配有噬菌体特异性 RNA 的新核糖体产生。

为了区分新旧核糖体，研究者用重同位素^{15}N和^{13}C标记未感染细胞的核糖体，使“老”核糖体变重。然后，用 T2 噬菌体感染这些细胞并同时将它们转移到含有轻同位素^{14}N和^{12}C的培养基中。因此在噬菌体感染后产生的任何“新”核糖体都会变轻，并在密度梯度离心时与“老”的重核糖体分开。Brenner 及同事也用^{32}P标记感染的细胞以便给合成的任一噬菌体 RNA 做标记，然后提出问题：放射性标记的噬菌体 RNA 与“新”的还是“老”的核糖体结合？

图 3.12 显示在“老”核糖体上发现了噬菌体 RNA，而“老”核糖体 RNA 在感染之前就产生了。很明显，“老”rRNA 不能携带噬菌体的遗传信息。延伸来讲，它也不可能携带宿主的遗传信息。因此，核糖体是不变的，由其合成的多肽的性质取决于与其结合的 mRNA，这种关系就像 CD 机与 CD。音乐（多肽）的本质依赖于 CD（mRNA），而不是 CD 机（核糖体）。

其他研究者也已鉴定出一种更好的信使：一组与核糖体瞬时结合的不稳定 RNA。有趣的是，在 T2 噬菌体感染细胞中，这种 RNA 的碱基组成与噬菌体 DNA 的非常相似，而与细菌的 DNA 和 RNA 完全不同。这正是我们所期望的噬菌体**信使 RNA**（mRNA）。此外，与宿主 rRNA 不同，宿主 mRNA 碱基的组成与宿主的 DNA 相似。这进一步支持了 mRNA 是信息分子，rRNA 不是信息分子的假设。

小结 信使 RNA 将遗传信息从基因带到合成多肽的核糖体上。

转录

如所预期，转录遵循与 DNA 复制相同的碱基配对原则：DNA 中的 T、G、C、A 分别与 RNA 产物中的 A、C、G、U 配对（注意 RNA 中出现的尿嘧啶代替了 DNA 中的胸腺嘧啶）。这种碱基配对模式保证了 RNA 转录物是基因的忠实拷贝（图 3.13）。

当然，像转录这样高度定向的化学反应不可能通过自身完成——它们是酶催化的。指导转录的酶称为**RNA 聚合酶**（RNA polymerase）。图 3.14 是大肠杆菌 RNA 聚合酶的工作示意图。转录有 3 个阶段：起始、延伸和终止。以下是细菌转录的 3 个阶段。

1. 起始　首先，酶识别一段称为**启动子**（promoter）的区域，启动子就位于基因的“上游”。聚合酶紧紧地结合在启动子上并引起启动子内两条 DNA 链局部至少有 12bp 解链。接着，聚合酶开始合成 RNA 链，其作用底物是 4 种**核糖核苷三磷酸**（ribonucleoside triphosphate）：ATP、GTP、CTP 和 UTP，起始底物通常是嘌呤核苷酸。当第一个核苷酸进入合适的位置后，聚合酶会把第二个核苷酸加在第一个的后面，形成 RNA 链中的初始磷酸二酯键。在聚合酶离开启动子开始延伸之前会加上几个核苷酸。

2. 延伸　在转录的延伸阶段，RNA 聚合酶指导核糖核苷酸连续从 5′→3′方向（从 RNA 的 5′端到 3′端）加入到伸长的 RNA 链上。RNA 聚合酶沿 DNA 模板移动，“泡”状的解链 DNA 随其移动。在解链区 DNA 模板的碱基逐个暴露出来与新加入的核苷酸配对。转录过后，两条 DNA 链又彼此相互缠绕，重新形成双螺旋。转录与 DNA 复制的两个不

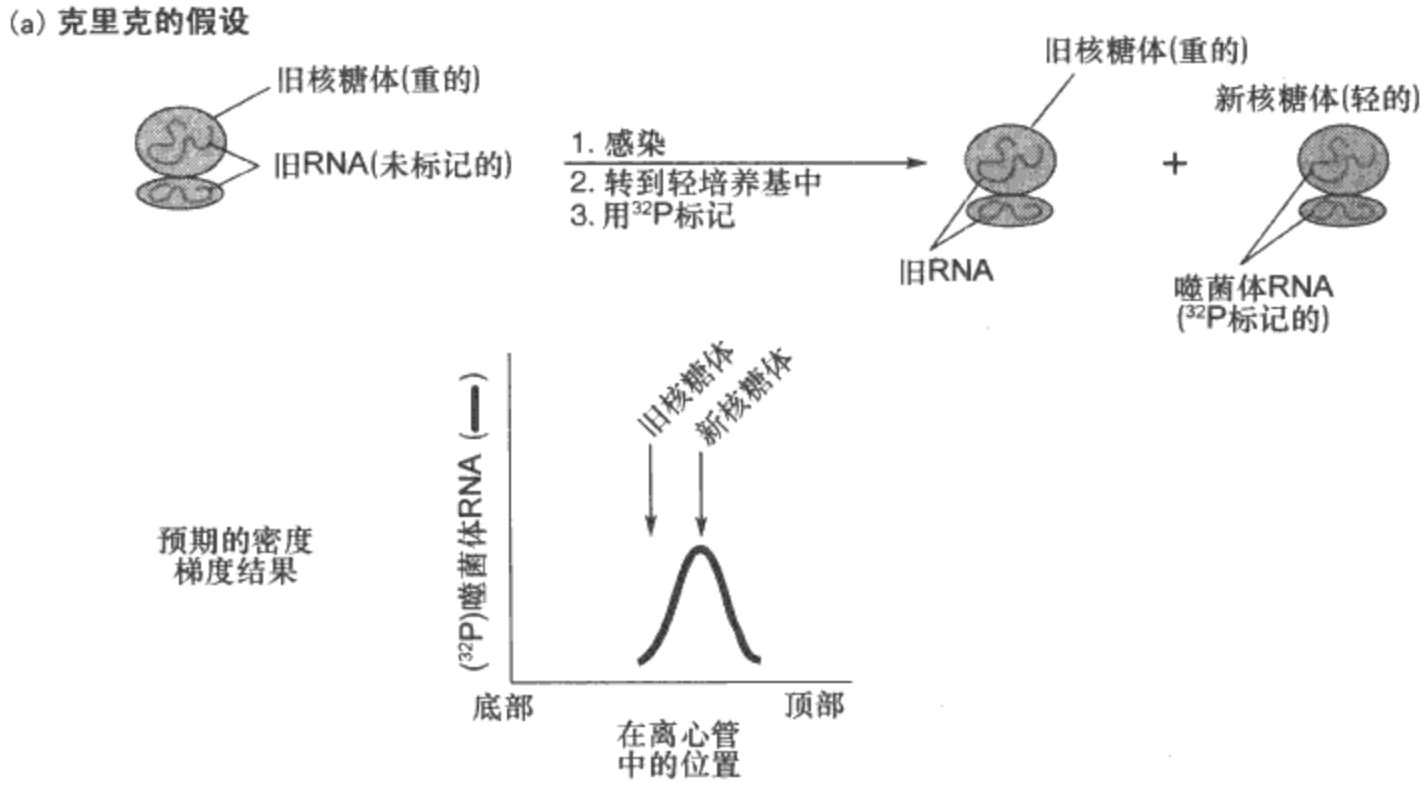


图 3.12　信使假说的实验证明。用碳和氮的重同位素标记细菌细胞，得到 *E. coli* 的重核糖体。然后用 T2 噬菌体感染细菌，并同时将其转移到含碳和氮正常同位素的“轻”培养基中培养，“轻”培养基中还加有^{32}P，以便使噬菌体 RNA 具有放射性。（a）克里克认为核糖体 RNA 携带合成蛋白质的信息。如果是这样，在噬菌体感染之后，就会产生带有噬菌体特异性核糖体 RNA 的整个新核糖体。那么，^{32}P 标记的新 RNA 将与新的、轻的核糖体（红色）一起移动。（b）Jacob 及同事提出有一个信使 RNA 携带遗传信息到核糖体。根据这个假设，噬菌体感染会产生新的被^{32}P 标记的噬菌体特异性信使 RNA（绿色），它们会跟旧的、重核糖体（蓝色）结合。因此，在密度梯度中放射性标记会跟旧的、重核糖体一起移动。这正是实际发生的情况。

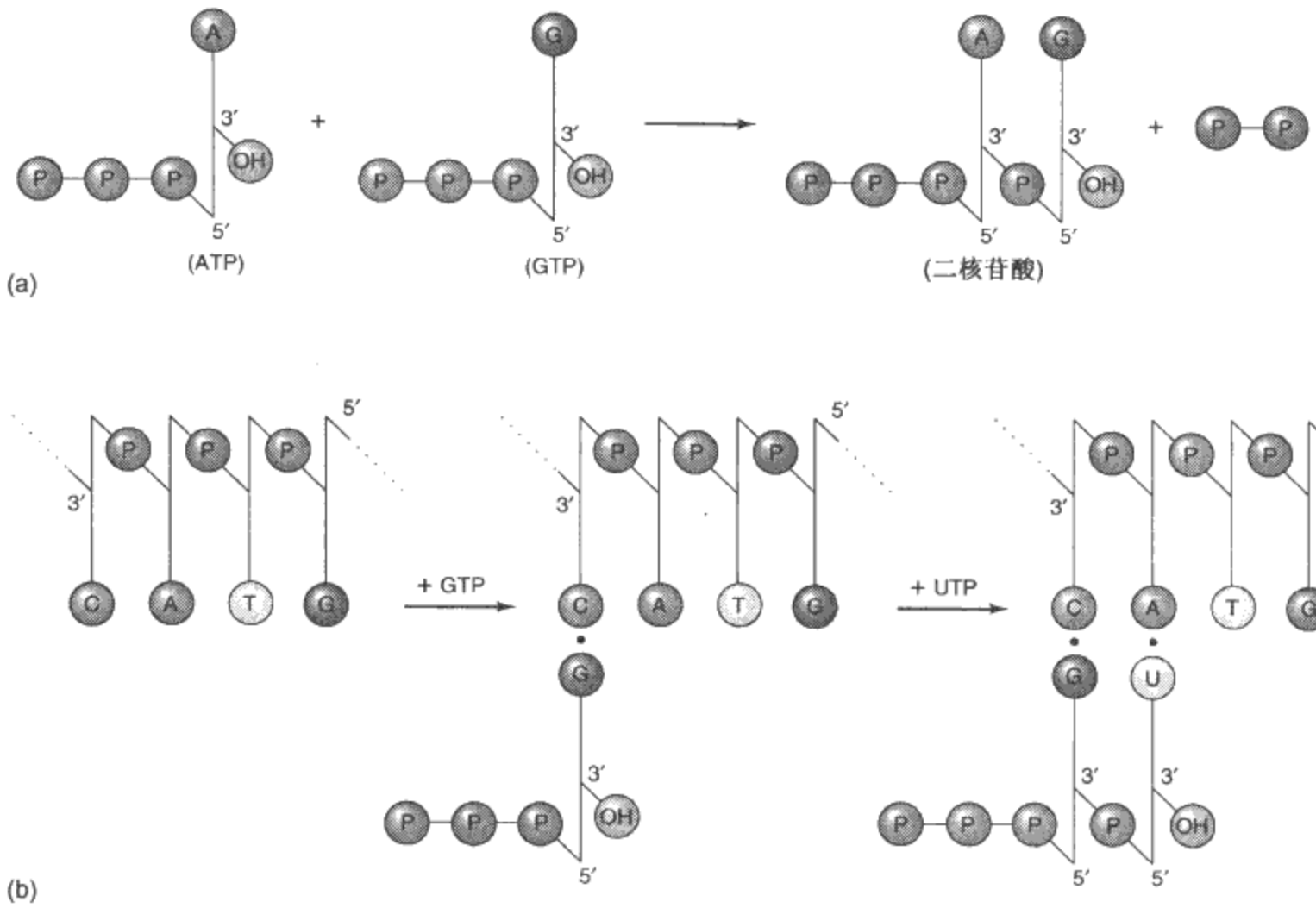


图 3.13　RNA 的合成。（a）RNA 合成中磷酸二酯键的形成。ATP 与 GTP 连接形成二核苷酸。离鸟嘌呤核苷酸最近的磷原子留在磷酸二酯键中，其余两个磷原子作为副产物焦磷酸被去除。（b）RNA 在 DNA 模板上的合成。上方的 DNA 模板含有序列 3′-dC-dA-dT-dG-5′。GTP 与 DNA 模板的 dC 核苷酸形成碱基对开始 RNA 合成。接着，UTP 提供一个尿嘧啶核苷酸与 DNA 模板上的 dA 碱基配对，并与 GTP 形成磷酸二酯键产生二核苷酸 GU。以同样方式，每一步有一个新核苷酸加入增长的 RNA 链上，直到转录完成。未显示副产物焦磷酸。

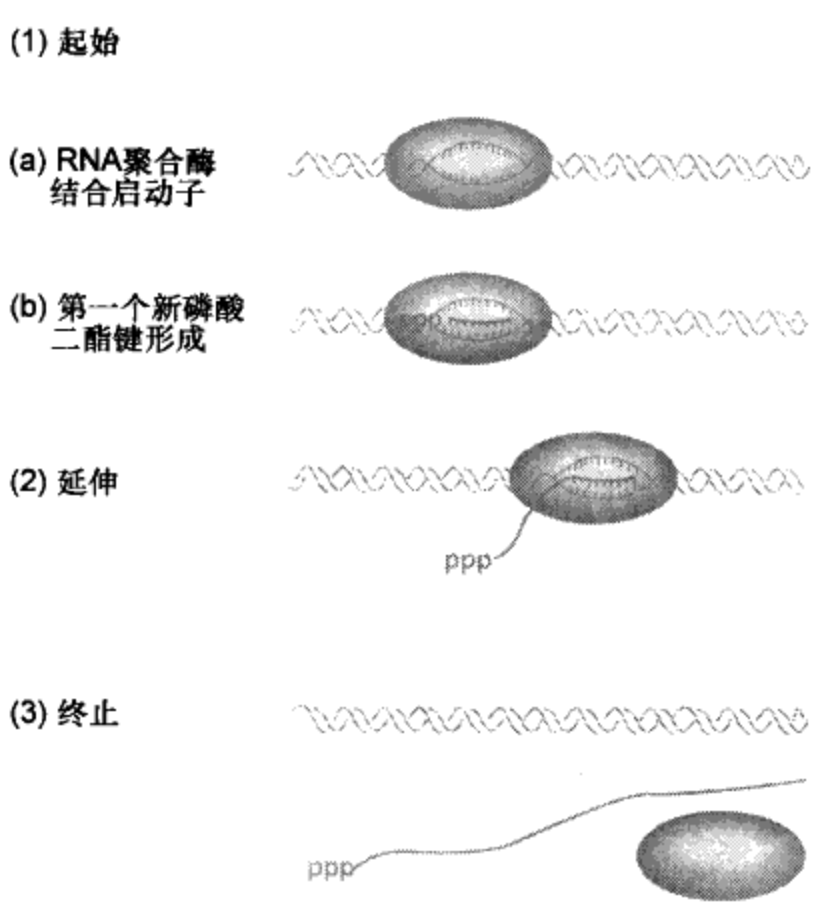


图 3.14　转录。（1a）起始的第一阶段，RNA 聚合酶紧密地结合在启动子上并解离一小段 DNA。（1b）起始的第二阶段，聚合酶通过磷酸二酯键将新生 RNA 的几个最早的核苷酸连一起。第一个核苷酸保留其三磷酸基团。（2）延伸过程中，双链 DNA 的解离泡随着聚合酶移动，使酶“阅读”DNA 模板链的碱基产生互补 RNA。（3）当聚合酶到达终止信号时终止发生，使 RNA 和聚合酶离开 DNA 模板。

同点：①RNA 聚合酶在转录时只生成一条 RNA 链，意思是在一个基因中它只复制一条 DNA 链（但另一条链可能会在另一个基因中被转录）。因此，转录是**不对称的**（asymmetrical）。而在 DNA 的半保留复制中，两条 DNA 都被复制了。②在转录中，DNA 的解链是有限和短暂的，只有出现足够长的链分离时，聚合酶才能“读取”DNA 模板。而在复制过程中，两条 DNA 母链永久性地分开。

3. 终止　如同启动子作为转录的开始信号一样，基因末端的区域称为**终止子**（terminator）。终止子与 RNA 聚合酶一同作用，松开 RNA 产物与 DNA 模板的结合，使 RNA 脱离 RNA 聚合酶和 DNA，转录终止。

最后，需要强调说明一些惯例：RNA 序列通常是从 5′端到 3′端的，即从左到右。因为 RNA 是沿 5′→3′方向生成的，mRNA 同样也是沿着 5′→3′方向翻译的。因为核糖体从 5′→3′方向阅读信息，因而从 5′→3′写 RNA 序列是合适的，以便我们像读句子一样读它。

基因同样习惯性地按它们从左到右的转录顺序书写。从一端到另一端的转录“流程”产生了“上游”，指 DNA 接近转录起始位置的区域（基因按习惯性书写时，指左端部分）。因此，我们可以说大部分启动子刚好位于它们各自基因的上游。同样，我们说基因通常位于启动子的下游。书写基因时，还习惯性地将其非模板链写在上方。

小结　转录以起始、延伸和终止 3 个步骤发生。起始包括 RNA 聚合酶结合到启动子上，局部解链及形成最初的磷酸二酯键。在延伸过程中，RNA 聚合酶沿着 5′→3′方向将核糖核苷酸连接在一起延伸 RNA。最后，在终止时聚合酶和 RNA 产物脱离 DNA 模板。

问题解答

问题　下面是 DNA 片段模板链的序列：

5′-ATGCGTGACTAATTCG-3′

a. 以习惯性双链形式写出这个 DNA 序列。

b. 假设其转录是从第一个核苷酸开始到最后一个核苷酸结束，以习惯方式写出转录物序列（5′→3′）。

解答

a. 模板链已经按 5′→3′方向给出了，但是模板链的习惯写法是 3′→5′，所以颠倒如下：

3′-GCTTAATCAGTGCGTA-5′

现在按照沃森-克里克碱基配对原则，加入非模板链，非模板链习惯性地写在上方：

5′-CGAATTAGTCACGCAT-3′

3′-GCTTAATCAGTGCGTA-5′

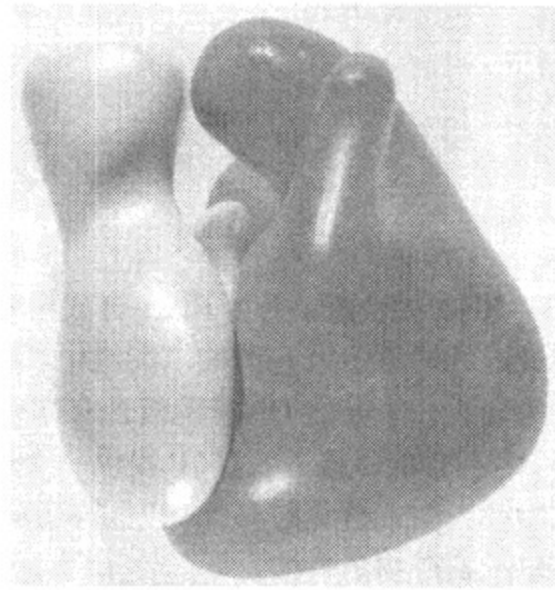
b. 已知基因的转录物与非模板链的序列且方向相同（用 U 代替 T），仅需要写出非模板链序列，并用 T 代替 U 即可

5′-CGAAUUAGUCACGCAU′-3′

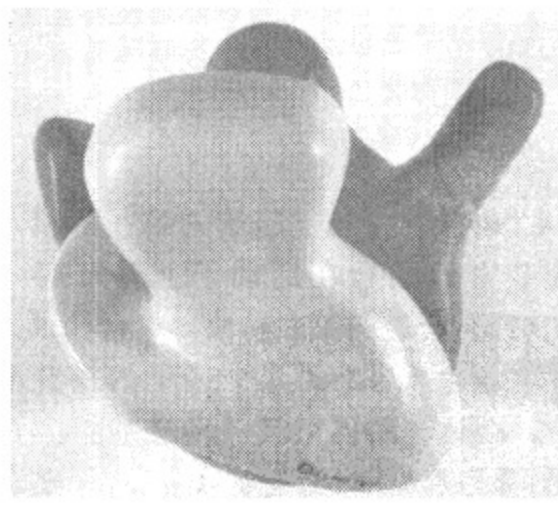
翻译

翻译的机制同样复杂而令人着迷，我们将在以后的章节讨论其细节。现在简要地看一下在翻译中起关键作用的两种物质：核糖体和转运 RNA。

核糖体：蛋白质合成的机器　图 3.15 是 *E. coli* 核糖体的大概形状及其 50S 和 30S 两个亚基，亚基以其沉降系数命名。沉降系数是溶液在超速离心时颗粒沉降速度的测量值。50S 亚基以其较大的沉降系数在离心力作用下朝离心管底部移动得较快。沉降系数是颗粒质量和形状的函数。重颗粒比轻颗粒



(a)



(b)

图 3.15　*E. coli* 核糖体的结构。（a）70S 核糖体的侧面观，30S 颗粒和 50S 颗粒组合在一起。（b）将（a）中的 70S 核糖体旋转 90°。30S 颗粒在前面，50S 颗粒在后面。（*Source*: Lake, J. Ribosome structure determined by electron microcopy of *Escherichia coli* small subunits, large subunits, and monomeric ribosomes. *J. Mol. Biol.* 105 (1976), p. 155, fig. 14, by permission of Academic Press.）

沉降得快，球状颗粒比长形或扁平颗粒移动得快，就像跳伞运动员下降时手脚缩拢要比手脚伸展下降得快。50S 亚基的质量约为 30S 的 2 倍，两者合在一起组成 70S 的核糖体。注意数字不是直接相加的，沉降系数与颗粒重量不呈比例，基本上是颗粒重量的 2/3 次方。

每个核糖体都含有 RNA 和蛋白质。30S 亚基由 1 分子 16S 的**核糖体 RNA**（ribosomal RNA，rRNA）加上 21 个核糖体蛋白组成。50S 亚基由 2 个 rRNA（23S＋5S）和 34 个核糖体蛋白组成（图 3.16）。所有这些核糖体蛋白自身当然是基因的产物，因此核糖体由数十个不同的基因产生。真核生物的核糖体甚至更复杂，有一个以上的 rRNA 和更多的蛋白质。

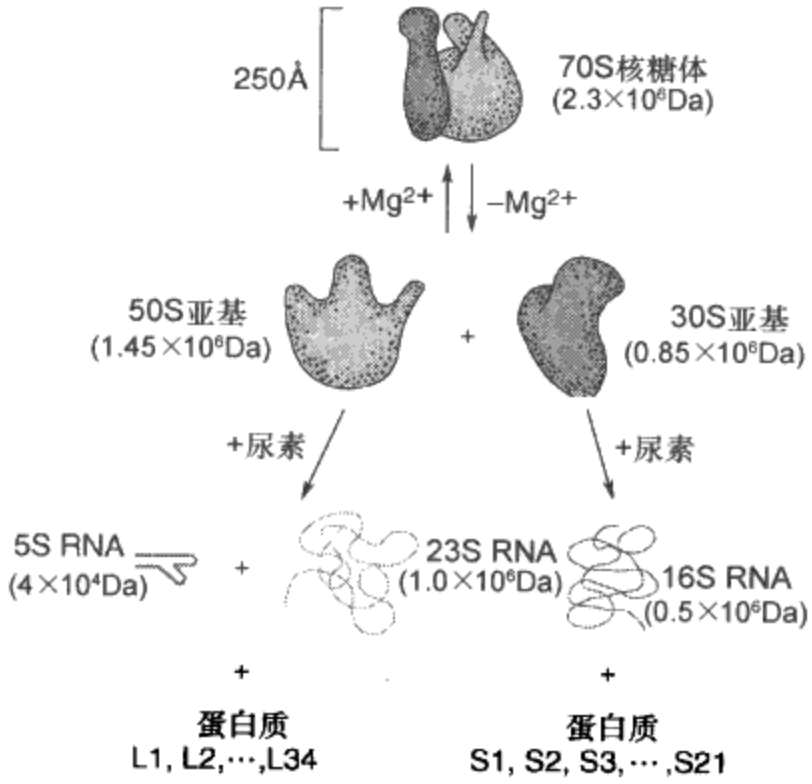


图 3.16　*E. coli* 核糖体的组成。图顶端的箭头表示去除镁离子时，70S 核糖体解离成两个亚基。下面的箭头表示在蛋白质变性剂的作用下，每个亚基解离成 RNA 和蛋白质。核糖体及其组分的分子质量（M_r，道尔顿表示）在圆括号中给出。

注意，rRNA 参与蛋白质的合成但不编码蛋白质，转录是 rRNA 基因表达的唯一步骤（除了对转录物的修饰），这些 RNA 不会发生翻译。

小结　核糖体是细胞的蛋白质工厂。细菌的核糖体为 70S，由 50S 和 30S 两个亚基组成。每个亚基都含有 rRNA 和大量蛋白质。

转运 RNA：适配器分子（adapter molecule）

转录的机制很容易预测，RNA 也遵循同样的碱基配对原则。根据这些原则，RNA 聚合酶合成其转录基因的复本。但什么规则控制核糖体将 mRNA 翻译成蛋白质呢？这是一个真正的翻译问题，核酸语言必须被转化成蛋白质语言。

在有大量实验证据支持之前，克里克在 1958 年的一篇论文中就提出了这个问题的答案。克里克分析认为所需要的是某种适配器分子，既能识别 RNA 语言中的核苷酸，又能识别蛋白质语言中的氨基酸。他是对的。他甚至注意到一种未知功能的小 RNA 有可能起适配器作用。他又对了！虽然他在这篇文章中也有一些错误的猜测，但无损其重要性。基于这些猜测的创新性，克里克的想法启发了进一步的研究（有些研究来自克里克自己的实验室），并最终破解了翻译之谜。

适配器分子实际上就是一种既能识别 RNA 又能识别氨基酸、被称为**转运 RNA**（transfer RNA，tRNA）的小分子 RNA。图 3.17 是 tRNA 识别苯丙氨酸（Phe）的示意图。在第 19 章我们将详细讨论 tRNA 的结构和功能。现在，我们用三叶草模型（尽管与 tRNA 真实形状的相似性有限）显示 tRNA 分子具有两个“作用”末端。一端在模型的上部，结合氨基酸。因为这是苯丙氨酸特异性 tRNA（$tRNA^{Phe}$），只有苯丙氨酸可以结合，苯丙氨酸-tRNA合成酶催化这一反应。这类酶的一般叫法是**氨酰-tRNA 合成酶**（aminoacyl-tRNA synthetase）。

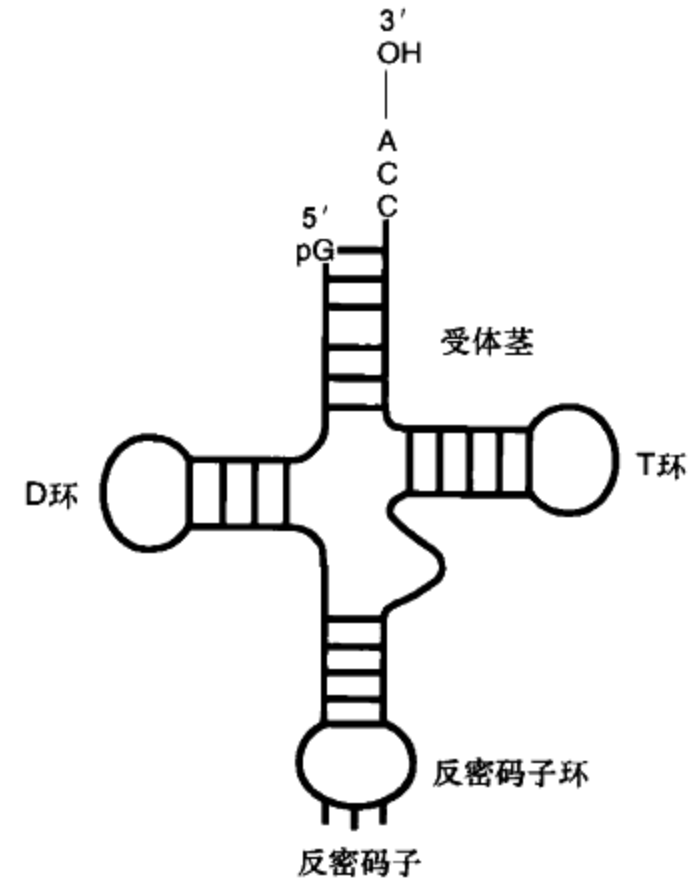


图 3.17　酵母 $tRNA^{Phe}$ 的三叶草形结构。顶端是受体茎，此处氨基酸与 3′端腺苷酸结合。左边是二氢尿嘧啶环（D 环），至少含有一个二氢尿嘧啶碱基。底端是含有反密码子的反密码子环。T 环（右边）含有几乎不变的 TψC 序列。每个环都被与其同一颜色的碱基配对臂所限定。

另一端在模型的底部，含有一个三碱基序列与 mRNA 的三碱基互补配对。mRNA 上的这种三联体

称为**密码子**（codon），其在 tRNA 上的互补体称为**反密码子**（anticodon）。本例中，mRNA 密码子已经吸引了带有苯丙氨酸 tRNA 的反密码子，即该密码子告诉核糖体将一个苯丙氨酸插入正在延伸的多肽中。由核糖体介导的密码子与反密码子间的识别，像任何其他双链核酸一样，也遵循沃森-克里克原则。前两个碱基对完全遵守，第三个碱基被允许有一定的自由度，见第 18 章。

从图 3.18 看出 UUC 是苯丙氨酸的密码子，这意味着**遗传密码**（genetic code）是 3 个字母的。我们按以下方法预测可能的 3bp 密码子：从 4 个不同碱基中一次取出 3 个的排列数目为 $4^3=64$，但是只有 20 个氨基酸。难道一些密码子没有被利用？实际上 3 个密码子（UAG、UAA、UGA）编码终止信号，而其他密码子编码特定氨基酸。这意味着大部分氨基酸有一个以上的密码子，所以遗传密码是**简并的**（degenerate）。第 18 章全面介绍了密码子系统及其被破译的过程。

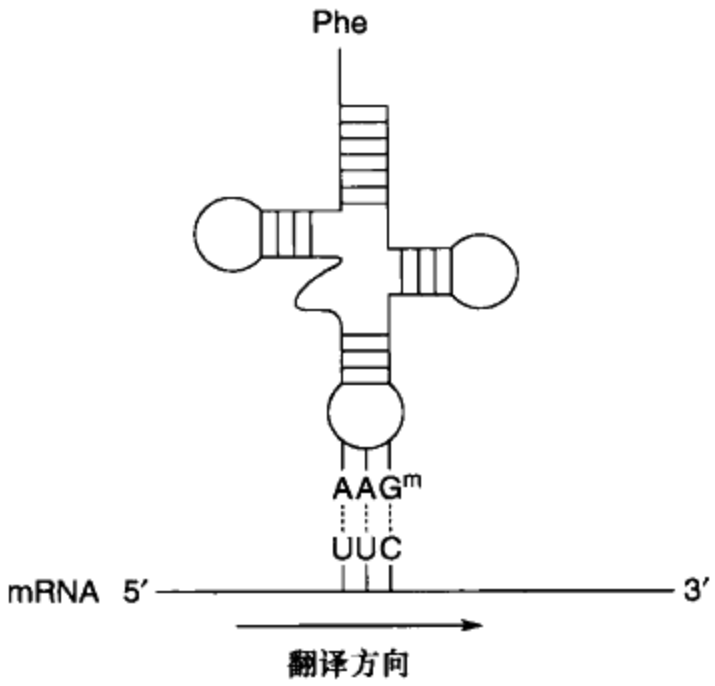


图 3.18　密码子-反密码子识别。mRNA 上的密码子与 tRNA 上相应反密码子之间的识别也遵循沃森-克里克规则。这里 $tRNA^{Phe}$ 上的 3′ AAG^m5′反密码子识别 mRNA 上的苯基丙氨酸密码子 5′UUC3′。G^m 表示甲基化的 G，它可以正常地碱基配对。注意：tRNA 相对于正常习惯（5′→3′）从左到右，是反向的（3′→5′），从而使反密码子以正确的方向（3′→5′，从左到右）与密码子配对。记住，DNA 的两条链反向平行，这适用于任何双链多聚核苷酸，包括小至 3bp 的密码子-反密码子对。

小结　tRNA 的两个重要位点使它们既能识别氨基酸又能识别核苷酸。一个位点共价结合一个氨基酸，另一个位点则含有与 mRNA 中 3bp 密码子配对的反密码子。因此，tRNA 能够作为克里克假设的适配器，并且是翻译机制的关键所在。

蛋白质合成的起始　我们刚才看到有 3 个密码子终止翻译，一个密码子（AUG）通常启动翻译。这两个过程的机制很不相同。在第 18 章将看到，3 个终止密码子与蛋白质因子相互作用，而起始密码子是与一个特殊的氨酰-tRNA 相互作用。在真核生物中这个特殊的氨酰-tRNA 是甲硫氨酰-tRNA（tRNA加载甲硫氨酸），而在原核生物是其衍生物——*N*-甲酰甲硫氨酰-tRNA，在甲硫氨酰-tRNA 的甲硫氨酸的氨基上连接一个甲酰基。

AUG 密码子不只位于 mRNA 的起始位置，也存在于中间。AUG 位于起始位置时，作为起始密码子，位于中间位置时，就编码甲硫氨酸，区别在于所处位置。细菌 mRNA 有一段特殊序列，以其发现者命名为 Shine-Dalgarno 序列，刚好位于起始密码子 AUG 的上游。Shine-Dalgarno 序列将核糖体吸引到附近的 AUG 上使翻译起始。真核细胞没有 Shine-Dalgarno 序列，而是在其 mRNA 的 5′端有一个特殊的甲基化核苷酸，叫做帽子。帽子结合蛋白 eIF4E 结合在帽子上帮助 mRNA 吸引核糖体。我们将在第 17 章详细讨论这一现象。

小结　AUG 通常是起始密码子，其与内部 AUG 的区别在于所处的位置不同。在细菌 mRNA 的起始端附近有 Shine-Dalgarno 核糖体结合序列，而在真核 mRNA 的 5′端有帽子结构。

翻译的延伸　在翻译起始阶段的末期，起始氨酰-tRNA 结合在核糖体上的 **P 位点**（P site）。核糖体需要给起始氨基酸一次加入一个氨基酸从而使延伸阶段开始。第 18 章将详细研究这个过程。在大肠杆菌中翻译的延伸总体过程如图 3.19 所示。延伸始于第二个氨酰-tRNA 与核糖体上 **A 位点**（A site）的结合，这个过程需要**延伸因子**（elongation factor）EF-Tu。EF 代表“延伸因子”，能量由 GTP 提供。

接着，在两个氨基酸之间必须形成肽键。核糖体大亚基含有一种**肽基转移酶**（peptidyl transferase），使 P 位点氨基酸或肽［在本例是甲酰甲硫氨酰.(fMet)］与 A 位点的氨酰-tRNA 的氨基酸部分形成肽键，结果在 A 位点产生一个二肽基-tRNA。二肽由 fMet 加第二个氨基酸形成，第二个氨基酸仍与其 tRNA 结合着。大核糖体 RNA 含有肽基转移酶的活性中心。

延伸的第三个阶段是**移位**（translocation），mRNA 在核糖体上移动一个密码子的长度。二肽基-tRNA 从 A 位点转移到 P 位点，去氨酰-tRNA 从 P 位点移至 E 位点（E site），由此提供一个离开核糖体的出口。移位需要另一个延伸因子 EF-G 和 GTP。

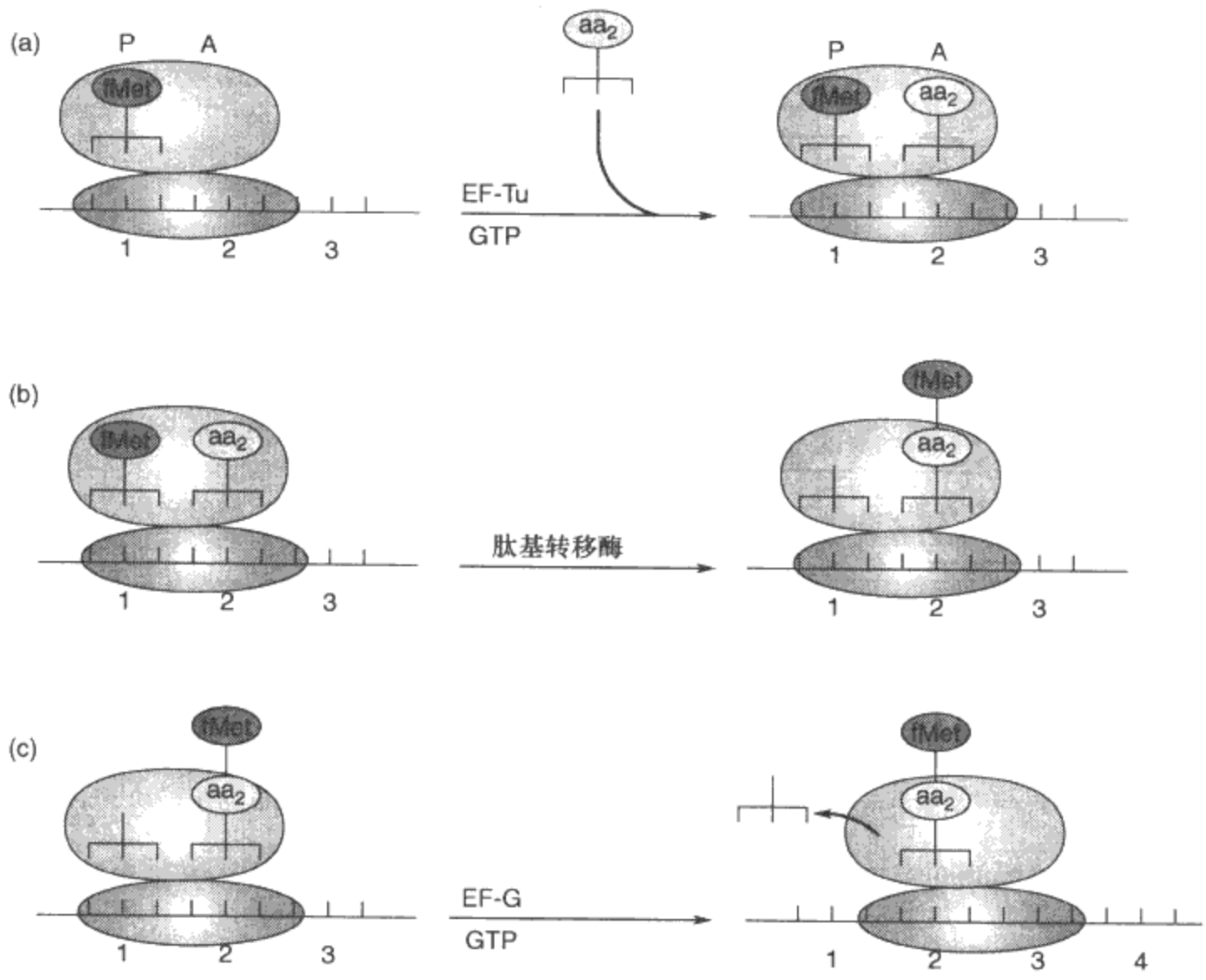


图 3.19　翻译延伸的概要。（a）EF-Tu 在 GTP 的帮助下将第二个氨酰-tRNA 转移到 A 位点。（按照惯例，P 位点和 A 位点如上方图所示分别标在核糖体的左半边和右半边）。（b）肽基转移酶是 50S 亚基中大 rRNA 的主要部分，在 fMet 和第二个氨酰-tRNA 之间形成肽键。由此在 A 位点产生一个二肽基-tRNA。（c）EF-G 在 GTP 的帮助下，将 mRNA 在核糖体上移动一个密码子长度，使密码子 2 与氨酰-tRNA 一起到达 P 位点，密码子 3 到达 A 位点。EF-G 同样使氨酰-tRNA 离开 P 位点移入 E 位点（没有显示），并从那里被排出。A 位点现在可以接受另一个氨酰-tRNA 开始另一轮延伸了。

小结　翻译延伸包括 3 个步骤：①第二个氨酰-tRNA 转移到 A 位点；②P 位点的第一个氨基酸与 A 位点的第二个氨酰-tRNA 形成肽键；③mRNA在核糖体上移动一个密码子的长度，将新形成的肽基-tRNA 带到 P 位点。

翻译的终止和 mRNA 的结构　3 个密码子（UAG、UAA、UGA）引起翻译的终止。称之为**释放因子**（release factor）的蛋白质识别这些**终止密码子**（termination codon 或 stop codon），使翻译停止同时释放多肽链。从基因编码区一端的起始密码子到另一端的终止密码子为一个**开放性可读框**（open reading frame，ORF）。之所以称为“开放”，是因为不含内部终止密码子，不会打断相应 mRNA 的翻译。“可读框”是指核糖体能以 3 种不同方式或“框”阅读 mRNA，采用哪种框取决于它从哪里开始。

图 3.20 显示了可读框的概念。这个小基因（比期望的任何基因都短）含有一个起始密码子（ATG）和一个终止密码子（TAG）（记住，这些 DNA 密码子会被转录成有相应密码子 AUG 和 UAG 的 mRNA）。在它们中间（包括这些密码子）有短的开放性可读框能被翻译成五肽（含有 5 个氨基酸的肽）：fMet-Gly-Tyr-Arg-Pro。理论上，翻译也可以从上游 4 个核苷酸处的另一个 AUG 开始，但是注意，此时翻译在另一个可读框进行，密码子也不同了，即 AUG CAU GGG AUA UAG。在这个可读框翻译产生四肽：fMet-His-Gly-Ile。第三个可读框没有起始密码子。天然 mRNA 可能具有一个以上的开放可读框，但通常使用最长的。

图 3.20 也显示该基因的转录和翻译在不同的地方开始和停止。转录开始于第一个 G，而翻译则始于下游 9bp 的起始密码子（AUG）处。因此，该基因的 mRNA 有一个 9bp 的**前导区**（leader），也叫做

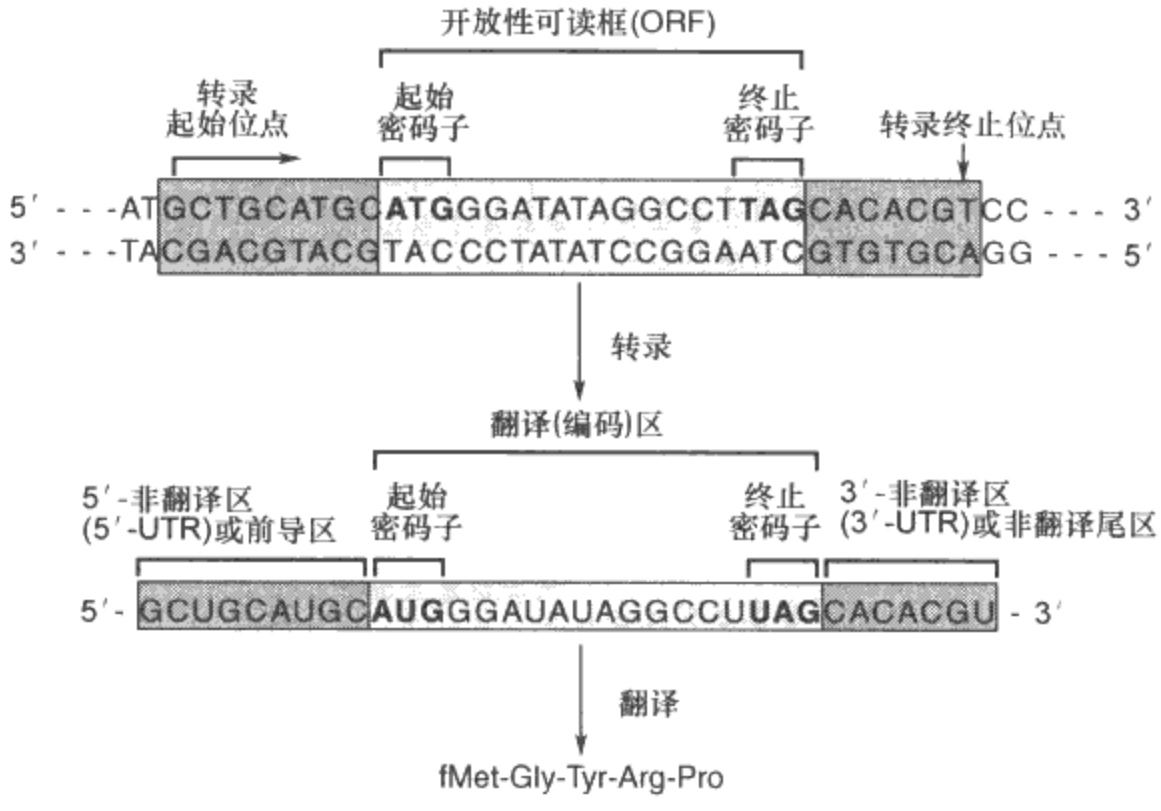


图 3.20　简化的基因和mRNA结构。上方是一个简化的基因，始于转录起始位点，止于转录终止位点。中间是翻译的起始密码子和终止密码子，由它们界定可被翻译产生多肽的开放性可读框（本例是一条非常短的只有 5 个氨基酸的多肽）。基因被转录产生一条具有编码区的 mRNA，始于起始密码子，止于终止密码子。这是基因中开放性可读框的 RNA 对等物。mRNA 中起始密码子上游的一段是引导序列或 5′非翻译区，终止密码子下游的一段是尾部或 3′非翻译区。注意，在上游向前第 4 个碱基处还有另一个开放性可读框，但它编码只有 4 个氨基酸的更短多肽。还要注意这个可读框相对于另一个可读框向左移动了 1bp。

5′-非翻译区（5′-untranslated region）或 **5′-UTR**。与之相似，在 mRNA 末端终止密码子和转录终止位点之间的存在一个**非翻译尾区**（trailer），也叫做 **3′-非翻译区**（3′-untranslated region）或 **3′-UTR**。在真核基因中，转录终止位点可能在下游很远处，但 mRNA 可以在终止密码子的下游剪切，一串 A［poly（A）］被加到 mRNA 的 3′端，这时非翻译尾区是终止密码子和 poly（A）之间的一段 RNA。

小结　翻译在终止密码子（UAG、UAA 或 UGA）处结束。开放性可读框包括翻译起始密码子、编码区和终止密码子。mRNA 5′端与起始密码子间的片段被称为前导区或 5′-UTR，3′端［或 poly（A）］与终止密码子间的片段被称为非翻译尾区或 3′-UTR。

3.2　复制

基因的第二个特征是忠实地复制。沃森-克里克模型（见第 2 章）设想 DNA 复制时新 DNA 链产生遵循 A-T、G-C 碱基配对原则。这一原则很关键，因为 DNA 复制机器必须能够区别配对的好坏。DNA 复制模型还显示两条母链分离，各自作为子链的模板，称为**半保留复制**（semiconservative replication），因为每个子代双链螺旋中有一条母链和一条新链（图 3.21a）。换句话说，在每个子代双螺旋中都“保留”一条母链。但这不是唯一的可能性。另一种可能机制（图 3.21b）是**保留机制**（conservative replication），两条母链不分开，以某种方式产生一个含有两条完整新链的子代螺旋。还有一种可能是**分散复制**（dispersive replication），DNA 变成片段，复制后新老 DNA 共存在于同一链中（图 3.21c）。正如第 1 章提到的，Matthew Meselson 和 Franklin Stahl 证明 DNA 的确是半保留复制。第 20

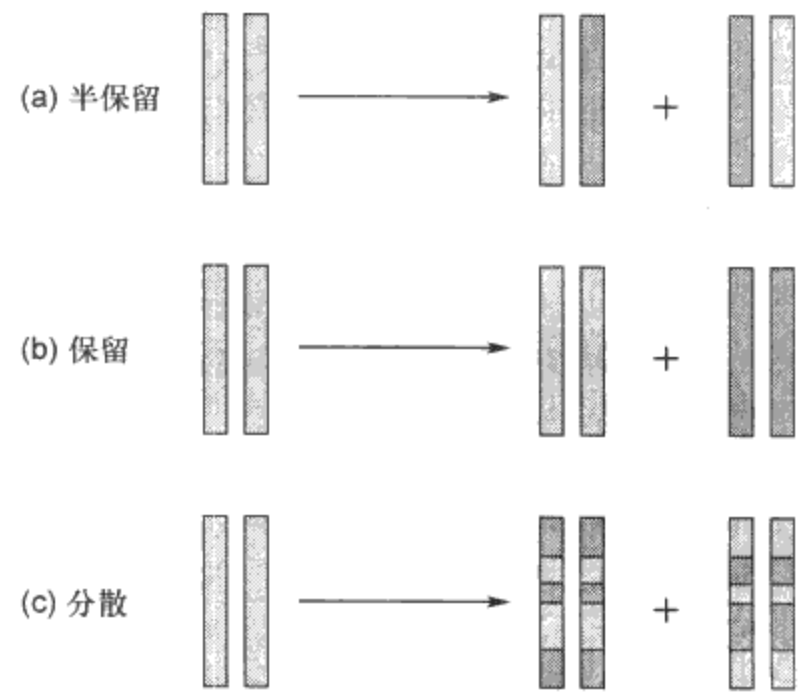


图 3.21　DNA 复制的 3 种假说。（a）半保留复制（也见图 2.15）产生两个子代双链 DNA，每个含有一条老链和一条新链；（b）保留复制产生两个子代双链 DNA，一个有两条老链（蓝色）另一个有两条新链（红色）；（c）分散复制产生两个子代双链 DNA，双链都是老 DNA 和新 DNA 的混合物。

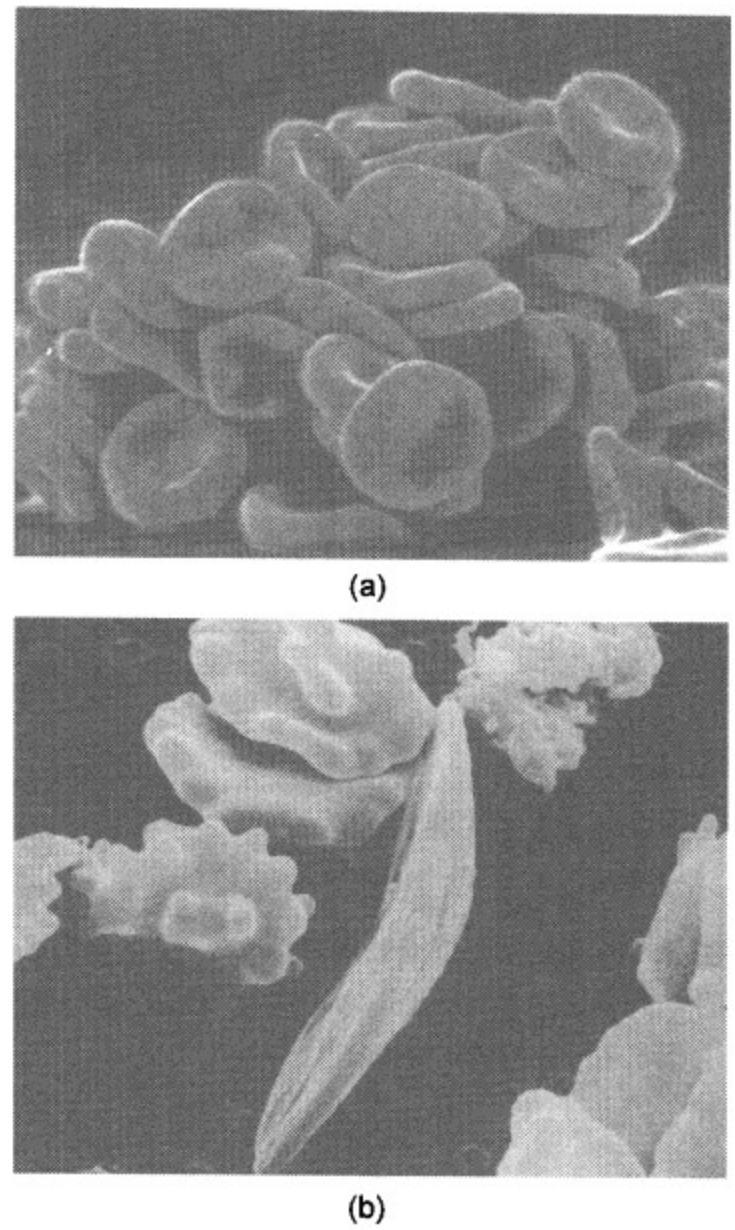
章将展示这个实验证据。

3.3 突变

基因的第三个特征是积累变化或突变。通过这一过程，生命本身发生变化，因为突变对进化至关重要。由于大部分基因是编码多肽的核苷酸链，而多肽又是氨基酸链，那么就很容易看出 DNA 中发生变化的结果。如果基因中某个核苷酸改变了，很可能那个基因蛋白质产物中的某个氨基酸将会发生相应改变。由于遗传密码的简并性，有时一个核苷酸的改变不会影响蛋白质。例如，密码子 AAA 突变为 AAG，两者都编码赖氨酸，所以检测不到，这种无危害的变化称为**沉默突变**（silent mutation）。更常见的是基因中核苷酸的改变导致蛋白质中氨基酸的相应改变。如果氨基酸的变化是保守的，则可能无害（如亮氨酸变成异亮氨酸），如果新氨基酸与原氨基酸差别很大，则会削弱或破坏蛋白质的功能。

镰状细胞贫血病

由一个缺陷基因引起疾病的典型例子是**镰状细胞病**（sickle cell disease），一种真正的遗传紊乱。纯合子患者的血液富含氧时，其红细胞外观正常。正常红细胞的形状是两面凹陷的圆盘状。然而，当患者做运动，或以其他方式耗尽血液中的氧时，红血细胞会迅速变成镰刀形或月牙形（图 3.22），这会产生严重后果。镰状细胞不易穿过毛细管，使毛细管堵塞和撑破，身体供血不足，引起内出血和疼痛。此外，镰状细胞非常脆弱，容易破裂，使患者贫血。不进行治疗的话，镰状细胞病患者会有致命危险。



(a)

(b)

图 3.22　正常红细胞和镰状细胞。扫描电镜显微照片比较了(a)双凹形正常细胞与(b)镰状细胞贫血病患者的变形红细胞。放大倍数：(a)3333×（b）5555×。(*Sources*：(a) ©Jeroboam/Omikron/Photo Researchers，Inc.（b）©Omikron/Photo Researchers，Inc.)

是什么导致红细胞变为镰刀形的呢？问题出在红细胞中红色的、运输氧的**血红蛋白**（hemoglobin）。正常血红蛋白在普通生理条件下是可溶性的，但是镰状细胞的血红蛋白在血液氧水平下降时会沉淀，形成长的、纤维状的聚积物使红细胞扭曲成镰刀形。

正常血红蛋白（HbA）与镰状细胞血红蛋白（HbS）之间有什么差异？1957 年 Vernon Ingram 采用 Frederick Sanger 发明的**蛋白质测序**（protein sequencing）方法测定了两种蛋白质的部分氨基酸序列，回答了这个问题。他重点研究了两种蛋白质的β-球蛋白。β-球蛋白是四聚体（4 链）血红蛋白的两种多肽链之一。先用酶选择性地打断肽键，将两条多肽切成片段。这些片段称为肽，能被叫做**指纹**（fingerprinting）（图 3.23）的二维技术分离开。用纸电泳做第一维分离，然后把纸旋转 90°做纸层析在第二维方向进一步分离肽。肽在纸上呈点状，不同的蛋白质因为氨基酸组成不同，产生不同的点状图形称为“指纹”。

Ingram 比较了 HbA 和 HbS 的指纹图谱，发现所有斑点都一致，只有一点例外（图 3.24）。该点在 HbS 指纹与 HbA 指纹有不同的迁移率，表明其氨基酸组成发生了改变。Ingram 检测了这两个点中两条肽链的氨基酸序列，发现它们是位于两个蛋白质开头处的氨基端肽，并且只有一个氨基酸不相同。HbA 第 6 位的谷氨酸在 HbS 中变成了缬氨酸（图 3.25）。这就是两个蛋白质中唯一的差别，但足以引起蛋白质行为的极度紊乱。

知道遗传密码后（第 18 章），我们要问β-球蛋白基因发生了什么变化导致蛋白质产物发生改变？

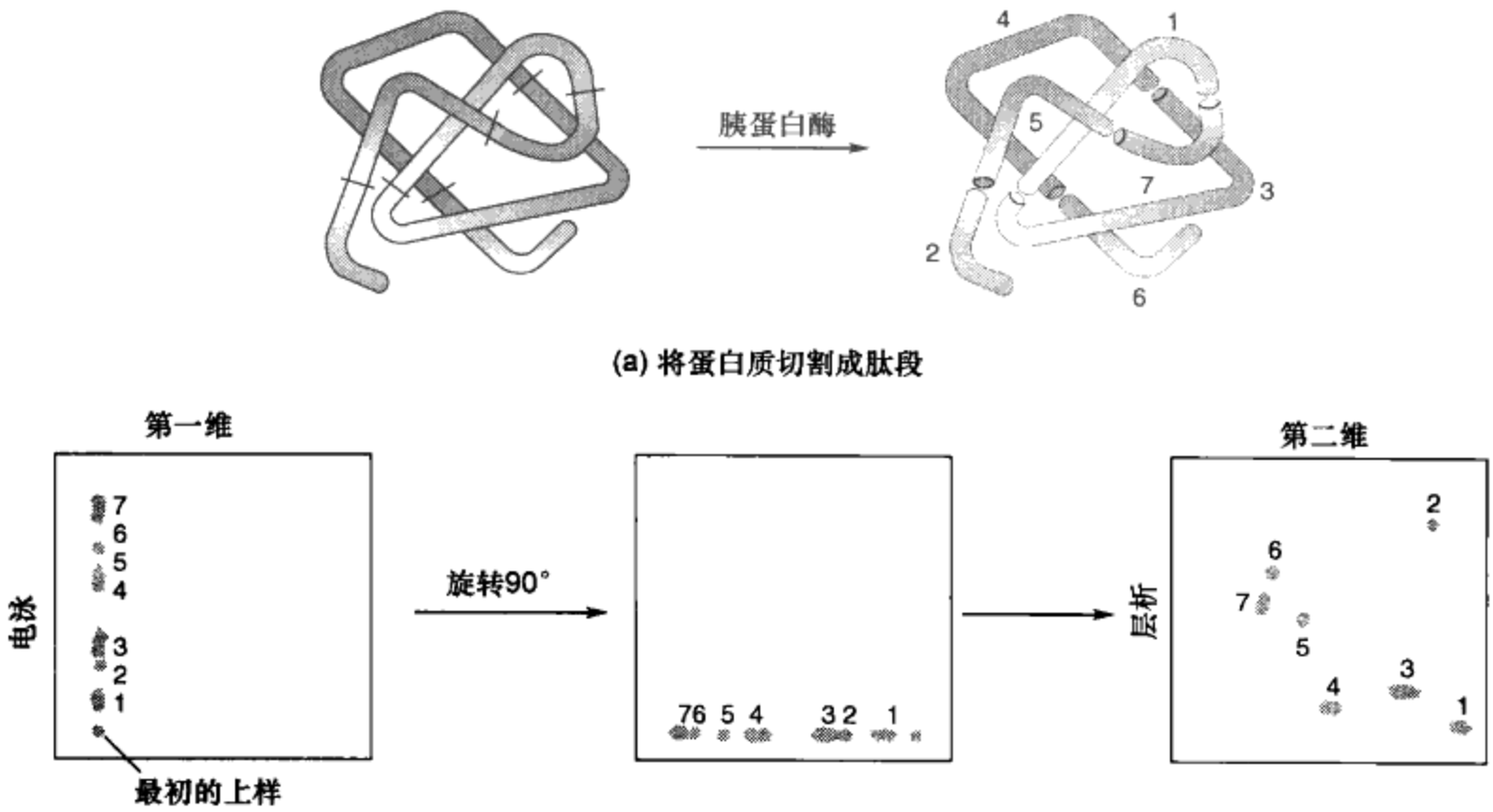


图 3.23　蛋白质的指纹图谱。（a）假想蛋白质有 6 个胰蛋白酶敏感位点（斜线表示）。用胰蛋白酶消化后，释放出 7 个肽段。（b）这些肽段在第一向电泳中部分分开，然后 90°旋转电泳纸，在第二向上再用层析法在另一种溶剂中完全分开。

图 3.24　血红蛋白 A 和血红蛋白 S 的指纹。除了圈住的肽之外，两者的指纹相同。该肽在 HbS 中向左上方移动了。（*Source*：Dr. Corrado Baglioni.）

谷氨酸（Glu）的 2 个密码子是 GAA 和 GAG，缬氨酸（Val）4 个密码子中的 2 个是 GUA 和 GUG。如

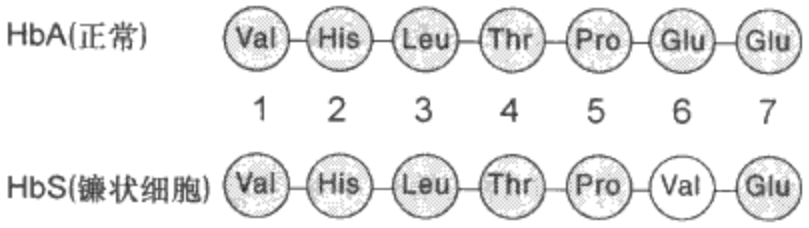


图 3.25　正常细胞和镰状细胞 β-球蛋白的氨基端肽序列。数字表示成熟蛋白质中相应氨基酸的位置。唯一区别是第 6 位由缬氨酸取代了 HbA 的谷氨酸。

果 HbA 基因中谷氨酸密码子 GAG 突变成 GTG，使 mRNA 变成 GUG，那么插入到 HbS 中的就是缬氨酸了。同样也可以认为是 GAA→GTA。注意，按照惯例一般表示的 DNA 链都是与 mRNA 有相同意义的（非模板链）。实际上，相反链（模板链）CAC，在 mRNA 中被转录成 GUG。图 3.26 总结了突变及其结果，可以看到基因如何变化引起产物的改变。

镰状细胞病在中非血统的人群中非常普遍。因此就自然产生了一个问题：为什么这种有害突变能够成功地在这个人群中传播呢？可以这样解释：虽然纯合子是致死性的，但杂合子没有，它们的正常等位基因可产生足够的产物防止血细胞形成镰刀形。而且，杂合子在中非还有一个优势：非洲的疟疾猖獗，HbS 能够帮助人类抵抗疟疾寄生虫的复制。

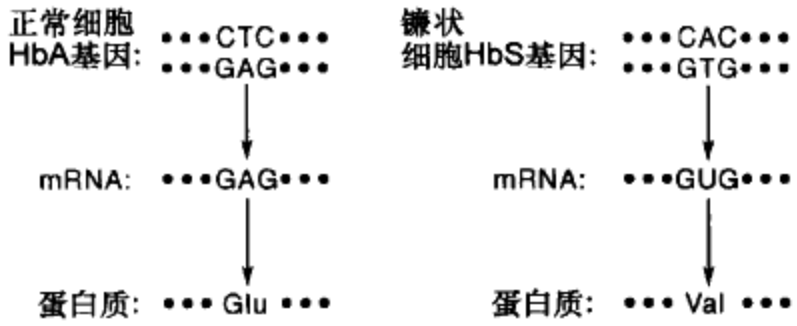


图 3.26　镰状细胞突变及其结果。正常基因非模板链的第 6 个密码子 GAG 变成了 GTG，导致镰状细胞 β-球蛋白 mRNA 上的第 6 位密码子从 GAG 变成了 GUG，进而导致镰状细胞 β-球蛋白第 6 位氨基酸由谷氨酸变成缬氨酸。

小结　镰状细胞病是一种人类遗传病，产生于 β-球蛋白基因的单碱基改变。改变的碱基引起 β-球蛋白某个位置插入了一个错误的氨基酸。在低氧条件下改变的蛋白质导致红血细胞的变形。这个疾病阐明了一个基本的遗传学概念，即基因的变化可导致其蛋白质产物的相应变化。

总结

基因的 3 种主要功能是信息储存、复制和积累突变。蛋白质或多肽是通过肽键连接的氨基酸多聚物。大部分基因含有合成一条多肽的信息，并在一个二步过程中表达：转录或基因的 mRNA 拷贝的合成，接下来将这个信息分子翻译成蛋白质。翻译在细胞的蛋白质工厂——核糖体的复杂结构中进行。翻译需要既能识别 mRNA 的遗传密码，也能识别 mRNA 编码的氨基酸的适配器分子。具有这两种能力的转运 RNA（tRNA）担负了这个任务。

翻译延伸包括 3 个步骤：①第二个氨酰-tRNA 转移到 A 位点；②P 位点的第一个氨基酸与 A 位点的第二个氨酰-tRNA 形成肽键；③mRNA 在核糖体上移动一个密码子的长度，将新形成的二肽基-tRNA带到 P 位点。翻译在终止密码子（UAG、UAA 或 UGA）处结束。RNA 或 DNA 的一个包括翻译起始密码子、编码区、终止密码子的区域称为一个开放性可读框。位于 5′端和起始密码子之间的 mRNA 片段叫做前导区或 5′-UTR，3′端［或 poly (A)］与终止密码子之间的部分被称为非翻译尾区或 3′-UTR。

DNA 以半保留方式复制，母链分离，各自作为模板合成新的互补链。基因中的变化或突变常能引起多肽产物中相应的位置发生变化。镰状细胞病就是这类突变产生危害的一个例子。

复习题

1. 图示氨基酸的一般结构。

2. 图示肽键的结构。

3. 用简图比较蛋白质的 α 螺旋和反向平行 β 折叠结构。为简化起见，只显示蛋白质的骨架原子。

4. 蛋白质的一级、二级、三级、四级结构各是什么意思？

5. 基于尿黑酸病症，Garrod 对蛋白质与基因间的关系有何预见？

6. 叙述 Beadle 和 Tatum 的实验方法，证明蛋白质与基因间的关系。

7. 基因表达的两个主要步骤是什么？

8. 叙述 Jacob 及同事证明 mRNA 存在的实验并给出结果。

9. 转录的 3 个步骤是什么？用图分别说明。

10. *E. coli* 核糖体有哪些 rRNA？每种 rRNA 分别属于哪个核糖体亚基？

11. 图示 tRNA 的三叶草形结构，指出连接氨基酸的位置和反密码子的位置。

12. tRNA 是如何行使 mRNA 的 3bp 密码子与蛋白质氨基酸之间的适配器功能的？

13. 解释基因中单碱基变化怎样导致 mRNA 翻译的提前终止。

14. 解释基因中部的单碱基缺失如何改变其可读框。

15. 解释基因的单碱基变化怎样导致其产物中单个氨基酸的变化，举例说明。

分析题

1. 这是一个细菌基因的部分序列：

5′GTATCGTATGCATGCATCGTGAC3′

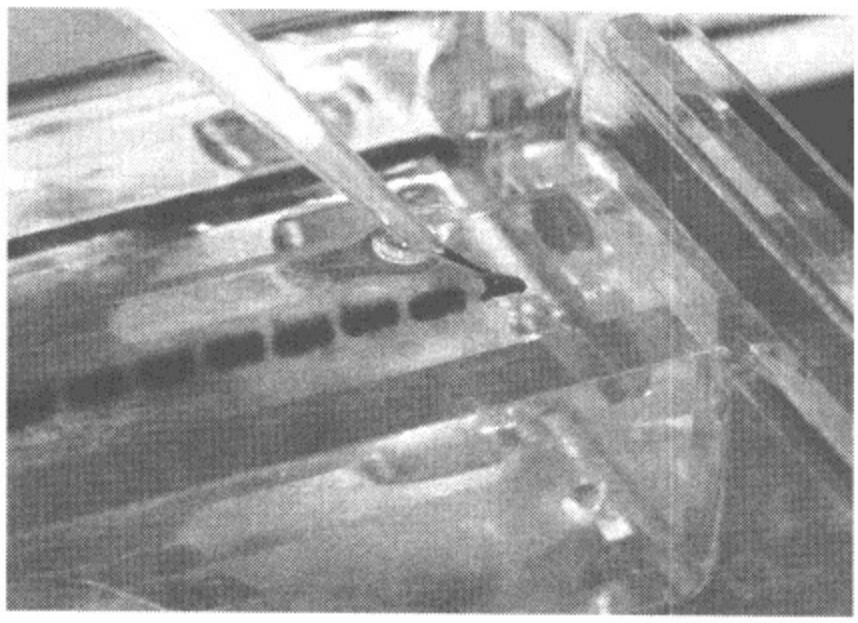
3′CATAGCATACGTACGTAGCACTG5′

模板链在下部，（a）假设转录始于模板链的第一个 T 并延续到末端，所产生的 mRNA 是什么？（b）找出这条 mRNA 的起始密码子。（c）模板链的第一个 G 变成 C 会对翻译有影响吗？如果有，是什么影响？（d）模板链的第二个 T 变成 G 会对翻译有影响吗？如果有，是什么影响？（e）模板链的第三个 T 变成 C 会对翻译有影响吗？如果有，是什么影响？（提示：无需知道遗传密码来回答这些问题，只需知道本章给出的起始密码子和终止密码子。）

2. 假设你用脉胞霉菌做类似于 Beadle 和 Tatum 的遗传学实验，分离得到一个泛酸盐缺陷型突变体，不能合成泛酸盐，需要提供泛酸才能生长。请问这一现象是因为泛酸途径的哪个步骤被阻断了？

翻译　马　纪　校译　张富春

第4章　分子克隆方法



DNA 电泳。将 DNA 样品加到琼脂糖凝胶中进行电泳分离。移液器头正将 DNA 样加入胶孔。 © *Klaus Guldbrandsen/SPL/Photo Researchers, Inc.*

在回顾了基因的基本结构和功能之后，就可以更详细地学习分子生物学了，重点是理解分子生物学家在阐明基因结构和功能研究中所做的实验。为此，我们暂不介绍那些主要的分子生物学实验技术。在第 4、5 章先介绍一些常用技术，其他技术穿插在书中有关章节加以介绍。我们先从使分子生物学发生革命性变化的基因克隆技术开始。

设想你是一位 1972 年的遗传学家，想从分子水平研究真核基因的功能。例如，你对人生长素基因（human growth hormone，hGH）的结构和功能感兴趣。那么该基因的序列是什么？其启动子是怎样的？RNA 聚合酶如何与该基因相互作用？基因发生了什么变化导致垂体功能减退性侏儒症（hypopituitary dwarfism）？

回答上述问题之前必须纯化出足够多（约 1mg）的基因来进行研究。1mg 听起来不多，但是想象一下要从人类基因组中纯化出来，这个量就惊人了。hGH 基因的 DNA 量远低于人类基因组 DNA 总量的百万分之一，即使能设法收集到这么大的量，也不知怎样从中分离出目的基因，总之你会身陷困境。

基因克隆技术很好地解决了这个问题。通过将真核基因与小的细菌 DNA 或噬菌体 DNA 连接起来，并将这些重组分子转入细菌宿主细胞，就可大量生产高质量的目的基因。让我们来看一下怎样在细菌和真核细胞中克隆基因。

4.1　基因克隆

克隆实验的一个产物就是一个**克隆**（clone），它是一组相同的细胞或组织。我们知道有些植物只需简单地扦插就可以被克隆（希腊语，*klon* 是枝条的意思），而有些则通过从植物中分离出单个细胞长成完整的植株而克隆，甚至脊椎动物也可被克隆。John Gurdon 将单一蛙胚的细胞核移植到多个去核的卵中得到了同一青蛙的多个克隆。1997 年名为多莉（Dolly）的绵羊在苏格兰克隆诞生，用的是成年母羊的乳腺上皮细胞和去核卵细胞。同卵双胞胎则是一种自然克隆。

基因克隆实验的一般程序是将外源基因放入细菌细胞中，然后分离出单个细胞，从每个单细胞培养出菌落（colony），菌落里的每个细胞都相同且都含有该基因。因此，只要保证外源基因能在宿主细胞内复制，就可通过克隆细菌（宿主菌）而克隆外源目的基因。Stanley Cohen、Herber Boyer 及同事在 1973 年进行了首例克隆实验。

限制性内切核酸酶的作用

Cohen 和 Boyer 的出色的实验依赖于极具价值的**限制性内切核酸酶**（restriction endonuclease）。20 世纪 60 年代后期 Stewart Linn 和 Werner Arber 在 *E. coli* 中发现了限制性内切核酸酶。这类酶的命名

是基于它们能将外源 DNA（如病毒 DNA）切碎而阻止其入侵，因此“限制了”病毒的宿主范围。此外，它们在外源基因内部切割，而不是从两端逐步消化，所以叫做内切核酸酶（endonuclease）（希腊语：*endo* 内部的）而不是外切核酸酶（exonuclease）（希腊语：*exo* 外部的）。Linn 和 Arber 希望他们的酶能在专一位点切割 DNA，就像精细打磨的分子刀一样来切削 DNA 片段。

由 Hamilton Smith 从流感嗜血菌（*Haemophilus influenzae*）Rd 菌株中分离出的一种酶在切割 DNA 时具有特异性。这个酶被命名为 *Hin*dII（发音为 Hin-dee-two），前 3 个字母取自产生该酶的微生物的拉丁名，头一个字母是属名的第一个字母，后面两个字母是种名的前两个字母（由 *Haemophilus influenzae* 得 *Hin*）。另外，有时还包括菌株名，如“d”来自 Rd 菌株。如果菌株仅产生一种限制酶，名字末端用罗马字母 I 表示，如果不止一种，则分别以罗马数字 II、III 表示，以此类推。

*Hin*dII 识别如下序列：

GTPy↓PuAC

CAPu↓PyTG

并在箭头所示位点切割 DNA 双链，其中 Py 表示嘧啶 T 或 C，Pu 表示嘌呤 A 或 G，而且只要这一序列出现，*Hin*dII 就切开 DNA。令分子生物学家高兴的是 *Hin*dII 仅仅是数百个具有各自特异性切割序列的限制酶中的一个。表 4.1 中列出了一些常用限制酶的来源及其识别序列。注意有些酶的识别序列仅为 4bp 而不是更常见的 6bp，因此其切割频率相对较高。因为 4bp 序列平均每 $4^4=256$bp 就会出现一次，而 6bp 的序列在 $4^6=4096$bp 出现一次。因此，6bp 的酶可将 DNA 切割成平均长度约为 4kb（kilobase）的片段。有些限制酶，如 *Not*I 识别 8bp 序列，其切割频率则更低（每 $4^8\approx65\,000$bp切割一次），因此称这些酶为**稀切酶**（rare cutter）。*Not*I 在哺乳动物 DNA 中的切割频率比预期的更低，其识别序列包含两拷贝的稀有二核苷酸 CG。此外还有一些酶，如 *Sma*I 和 *Xma*I 的识别序列相同但切割位点不同。这类可识别相同序列中不同位点的酶叫做**异裂酶**（heteroschizomer）（希腊文 *hetero* 意思是不同，*schizo* 意思是切割）或**新裂酶**（neoschizomer）（希腊文 *neo* 意思是新的）。将识别相同序列中相同位点的酶叫做**同裂酶**（isoschizomer）（希腊文 *iso* 意思是相等）。

限制酶的主要优点在于其在相同位点重复性切割 DNA 的能力。这一特性是分析基因及其表达的实验技术的基础。但这不是唯一的优点，许多限制酶在两条 DNA 链上产生交错切口（表 4.1 中偏离序列中心切割位点的酶），使链产生突出末端或**黏性末端**（sticky end），两个不同的 DNA 分子容易连接在一起。例如，*Eco*RI（发音为 Eeko-R-noe）酶切形成的黏性末端之间的互补序列

5′-G↓AATC-3′ → -G3′ + 5′AATTC-

3′-CTTAA↑G-5′ → -CTTAA5′ + 3′G-

表 4.1 常用限制性内切核酸酶的识别序列及作用位点

名称	识别序列*
*Alu*I	AG↓CT
*Bam*HI	G↓GATCC
*Bgl*II	A↓GATCT
*Cla*I	AT↓CGAT
*Eco*RI	G↓AATTC
*Hae*III	GG↓CC
*Hin*dII	GTPy↓PuAC
*Hin*dIII	A↓AGCTT
*Hpa*II	C↓CGG
*Kpn*I	GGTAC↓C
*Mbo*I	↓GATC
*Pst*I	CTGCA↓G
*Pvu*I	CGAT↓CG
*Sal*I	G↓TCGAC
*Sma*I	CCC↓GGG
*Xma*I	C↓CCGGG
*Not*I	GC↓GGCCGC

* 只显示 5′→3′左到右的一条 DNA 链，但限制性内切核酸酶实际切割双链 DNA，如文中的 *Eco*RI，箭头指示每个酶的切割位点。

还要注意，*Eco*RI 在识别序列的 5′端切割，产生 4 个碱基的突出端。*Pst*I 在其识别序列的 3′端切割，产生 3′黏性末端，而 *Sma*I 在其识别序列的正中切割，产生无碱基突出的平末端（blunt end）。

限制酶产生交错切口是因为其识别序列多呈二重对称性，即识别序列旋转 180°后还是同样的，其正读与倒读的结果一致。因此，如垂直箭头所示，*Eco*RI 可以在顶链的左侧 G 和 A 之间切割，也可以在底链的右侧 G 和 A 之间切割。

5′-G↓AATTC-3′，或

3′-CTTAA↑G-5′

具有二重对称性的序列也称为**回文结构**（palindrome）。回文就是正读和倒读都一样的句子，如拿破仑的悼词：“Able was I ere I saw Elba”或对意大利食品偏爱的一种表述：“Go hang a salami! I'm a lasagna hog.”都是典型的回文结构。虽然 DNA 回文结构正读、倒读的结果一样，但必须在两个方向都按 5′→3′读起，也就是说，读顶链时要从左向右，而读底链时要从右向左。

有关限制酶的最后问题是：如果能切割入侵的病毒 DNA，那么它们为什么不破坏宿主细胞自身的 DNA 呢？答案是几乎所有限制酶都有对应的甲基化酶，可以识别相同的位点并对其甲基化。

限制酶和甲基化酶统称为**限制-修饰系统**（restriction-modification system），或叫 **R-M 系统**。甲基化之后，DNA 的酶切位点被保护起来，宿主细胞中甲基化的 DNA 就不会受到损伤。但是 DNA 复制时怎么办？新合成的 DNA 链尚未甲基化，是否容易被切割呢？图 4.1 解释了 DNA 在复制过程中怎样继续受到保护。细胞 DNA 每次复制时，子代双链之一是新合成未甲基化的，而另一条链是已经甲基化的母链，这种半甲基化（hemimethylation）状态足以保护 DNA 双链不被多数限制酶所酶切，因而甲基化酶有时间寻找甲基化位点并使另一条链甲基化，产生完全甲基化的双链 DNA。

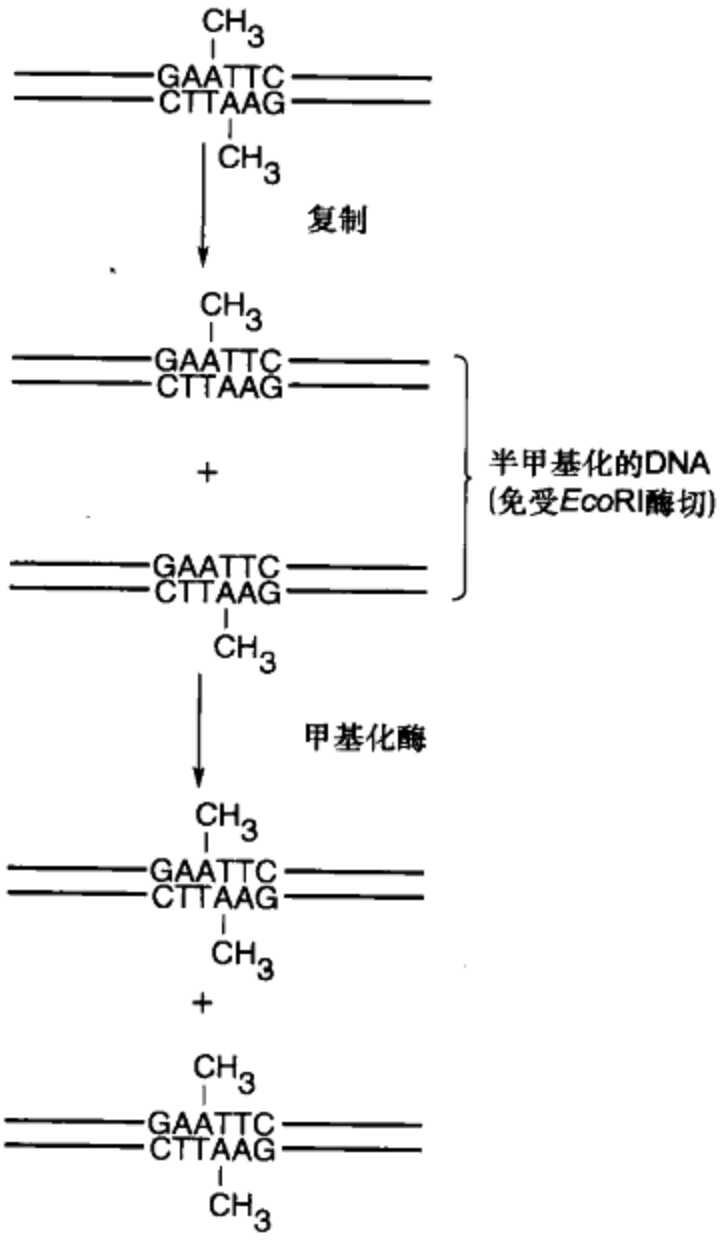


图 4.1　DNA 复制后保持对限制性内切核酸酶的抗性。以双链 DNA 的 *Eco*RI 位点的甲基化保护（CH_3）为例，转录后每个子代 DNA 双链中的母链仍保持甲基化状态，而另一条新合成的链没有被甲基化。双链 DNA 中一条链上的甲基化足以抵御 *Eco*RI 酶的降解。甲基化酶很快会识别未甲基化的 *Eco*RI 位点并对其甲基化，从而产生完全甲基化的 DNA 双链。

Cohen 和 Boyer 在其克隆实验中充分利用了限制酶产生黏性末端这一特点（图 4.2）。他们用 *Eco*RI 分别酶切两个不同 DNA。这两个 DNA 都是**质粒**（plasmid），即一类独立于宿主菌染色体之外的小环状 DNA。第一个质粒是 pSC101，带有四环素抗性基因。另一个是 RSF1010，具有抗链霉素和磺胺类的特性。两个质粒都仅有一个 *Eco*RI **限制位点**（restriction site），用 *Eco*RI 切割时变成带有相同

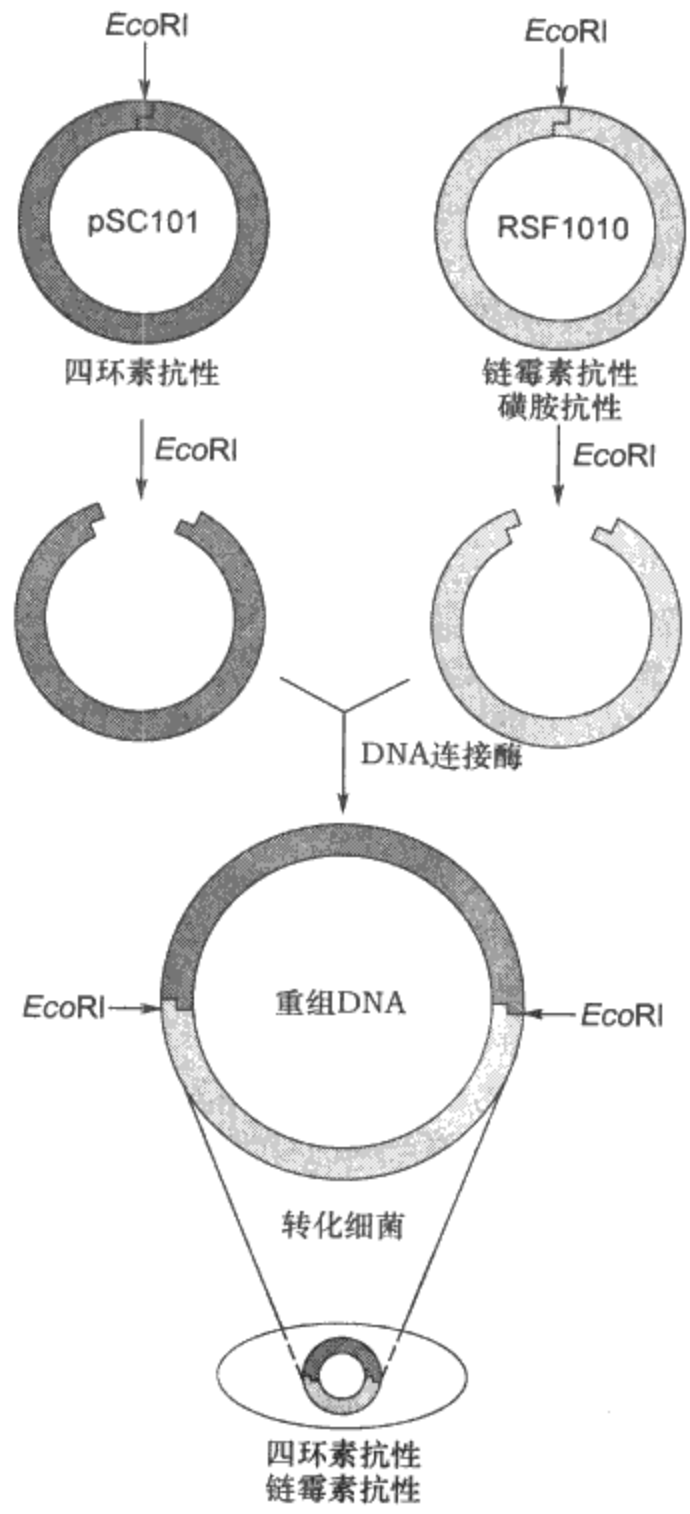


图 4.2　首次体外重组 DNA 的克隆实验。Boyer 和 Cohen 使用同一种限制性内切核酸酶处理两种质粒 pSC101 和 RSF1010，得到两条具有相同黏性末端的线性 DNA 片段，在体外可用 DNA 连接酶将其连接。通过转化将重组 DNA 再导入大肠杆菌细胞，筛选对四环素和链霉素都有抗性的克隆就是携带有重组质粒的克隆。

黏性末端的线状分子，这些黏性末端彼此通过碱基配对，至少能暂时连接在一起。当然，同一 DNA 分子黏性末端碱基配对会使 DNA 分子重新闭合成环状。最后，**DNA 连接酶**（DNA ligase）在两条 DNA 链末端间形成共价键，完成将两个 DNA 分子共价连接的任务。

该实验期望获得重组 DNA，即将原来独立的两段 DNA 连接在一起。虽然重组质粒在数量上可能远远少于切开后又重新自连接的原质粒，但是很容易筛选。连接产物重组质粒转化细菌细胞后赋予宿主两种抗性，pSC101 的四环素抗性和 RSF1010 的链霉素抗性。自然界有丰富的 DNA 重组现象，与其他大多数重组不同，该重组的 DNA 不是在细胞内自然形成的，而是分子生物学家在试管中连接起来的。

小结 限制性内切核酸酶识别 DNA 分子中特异序列，并在双链上进行切割。这种切割具有很强的特异性。同时由于在双链上产生的切口多是交错的，限制酶产生的黏性末端有助于将两个 DNA 连接起来，在体外形成重组 DNA。

载体

Cohen 和 Boyer 实验中所采用的两个质粒都能在 *E. coli* 中自我复制，因此两者都可作为载体使重组 DNA 复制。基因克隆实验都需要这种**载体**（vector），典型的基因克隆实验只需一种载体外加一段依赖载体进行复制的外源 DNA。外源 DNA 没有**复制起点**（origin of replication），必须将其置于带有复制起点的载体中，否则无法进行复制。自 20 世纪 70 年代中期以来，已开发出众多载体，主要分为两大类：质粒（plasmid）和噬菌体（phage）。

质粒载体 在研究克隆的早期阶段，Boyer 小组开发的 pBR 系列载体非常流行。现在这些早期的质粒已经不再使用了，鉴于它们提供了克隆方法的简单模型，在这里我们还是介绍一下。其中一个质粒是 **pBR322**（图 4.3），含有两个抗生素，氨苄青霉素和四环素的抗性基因，在这两个抗性基因之间有复制起点。该质粒含有几种常用限制酶的单个识别位点，如 *Eco*RI、*Bam*HI、*Pst*I、*Hin*dIII 和 *Sal*I 等。这种单一酶切位点非常方便，可用任一种酶产生外源 DNA 插入位点，而质粒 DNA 毫无损失。

例如，将外源 DNA 片段克隆到 pBR322 的 *Pst*I 位点（图 4.4）。首先，用 *Pst*I 分别酶切载体和外源 DNA，使之在载体和外源 DNA 片段上都产生 *Pst*I 黏性末端。接着，将酶切产物混合，并用 DNA 连接酶孵育。当载体和外源片段的相同黏性末端进行瞬时碱基配对时，DNA 连接酶就会封闭切口

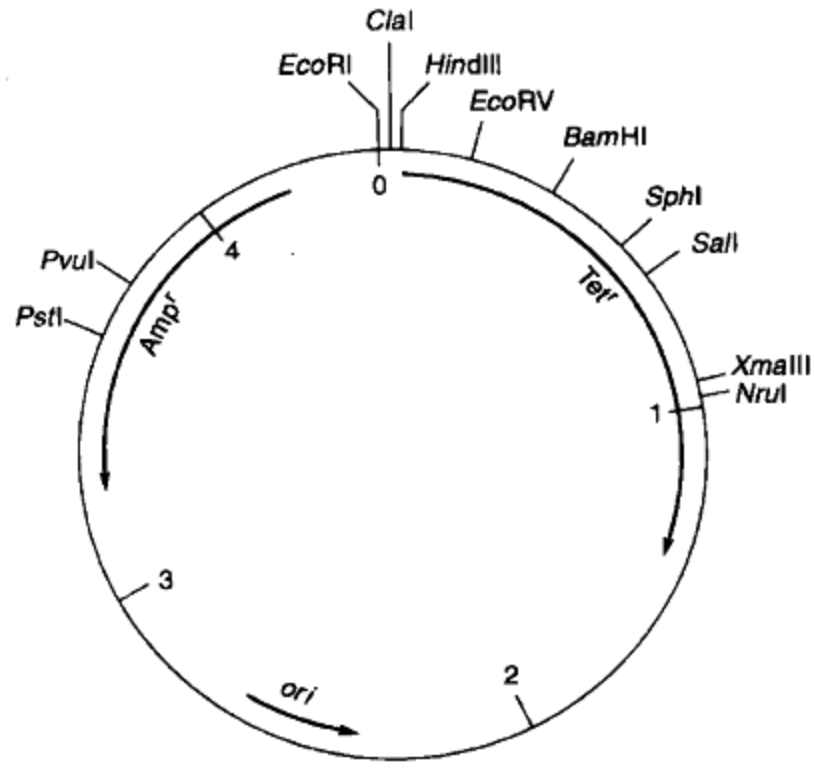
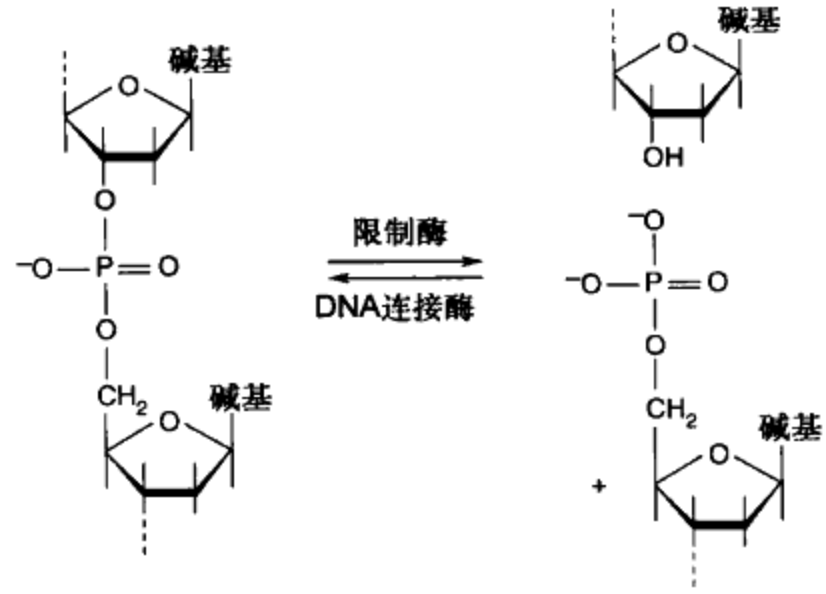


图 4.3 具有可用于插入外源 DNA 的 11 个独特限制位点的质粒 pBR322。两个抗生素抗性基因位点（Amp^r = 氨苄青霉素抗性；Tet^r = 四环素抗性）和复制起点（*ori*），数字是指以千碱基对（kb）表示的顺时针方向距离 *Eco*RI 位点的距离。



（nick），将这两个 DNA 片段共价连接起来。这个过程一旦完成，二者就不会再分开了，除非用 *Pst*I 重新酶切。切口只是磷酸二酯键的断裂，而 DNA 连接酶可重新形成磷酸二酯键，如下图所示：

下一步用 DNA 混合物转化 *E. coli*。传统方法是用浓缩的钙盐溶液孵育细菌细胞，使细胞膜出现缝隙，然后将这种具有通透性的细胞与 DNA 混合，使 DNA 进入其中。也可用高压电击将外源 DNA 导入宿主细胞，这种方法叫做**电穿孔法**（electroporation）。

如果所有酶切后的 DNA 都与质粒连接形成重组 DNA，那就简单了，但这永远不会发生。实际得到的是重新连接的质粒、重新连接的插入片段和重组 DNA 的混合物。那么如何区分呢？现在轮到载体上的抗生素基因发挥作用了。首先，在含有四环

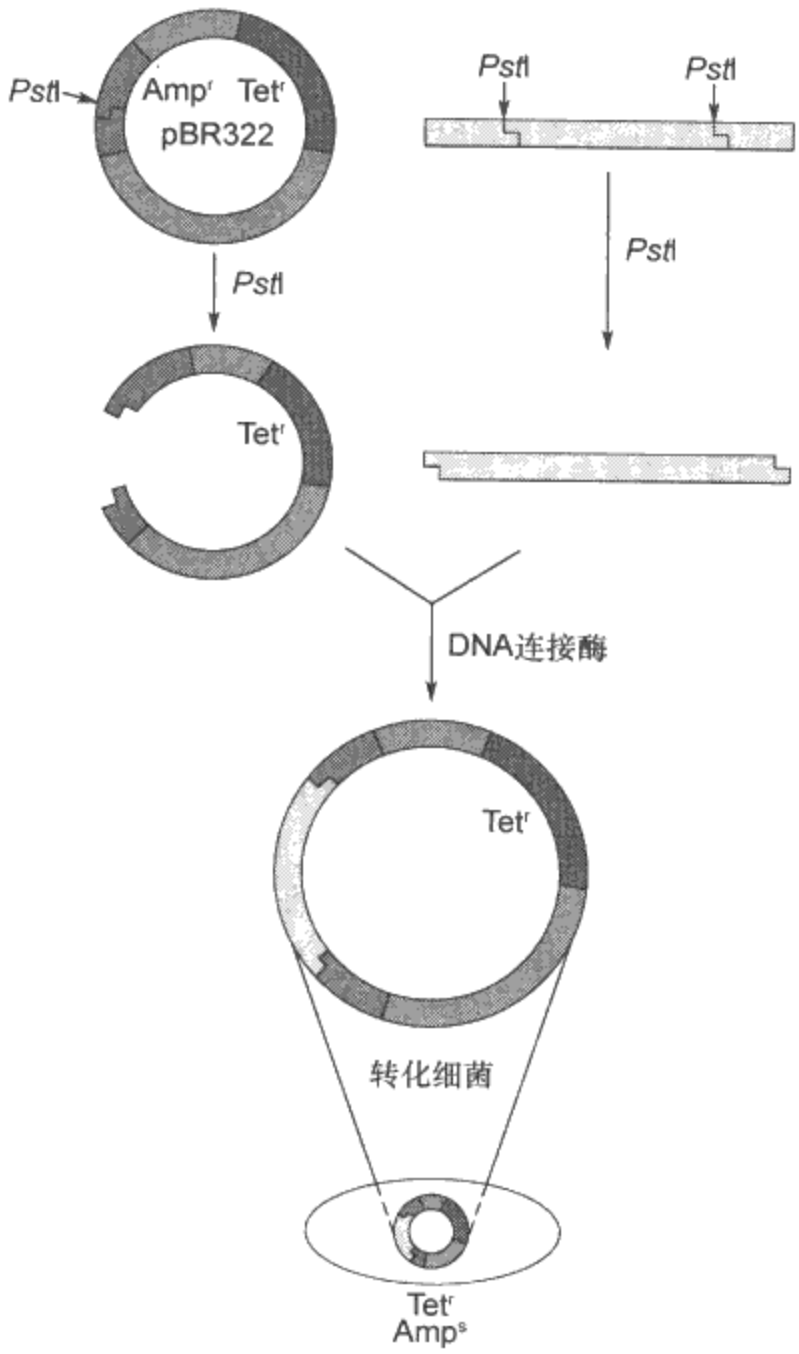


图 4.4　利用 pBR322 的 *Pst*I 位点克隆外源 DNA。用 *Pst*I 分别酶切质粒和插入 DNA，再用 DNA 连接酶连接其黏性末端。接着用重组 DNA 转化细菌并筛选四环素抗性和氨苄青霉素敏感性菌落。重组质粒因为外源基因中断了氨苄青霉素抗性基因而不再具有氨苄青霉素抗性。

素的培养上培养细菌细胞，筛选出含空载体或者含有外源 DNA 重组载体的细胞，而那些未转入 DNA 或仅转化了外源 DNA 的细菌则无四环素抗性而无法生长。

接下来要找出含有重组 DNA 的克隆。图 4.4 显示本实验中用于插入 DNA 片段的 *Pst*I 位点在氨苄青霉素抗性基因的内部。因此，外源基因插入到 *Pst*I 位点，会导致该基因的失活并使宿主菌表现出对氨苄青霉素的敏感性。据此，筛选那些四环素-抗性、氨苄青霉素-敏感的克隆就可找到含重组 DNA 的克隆。

怎样进行筛选呢？一种方法是**影印培养法**（replica plating）。将四环素抗性平板中菌落的拷贝转移到氨苄青霉素平板上，可以用图 4.5 所示的灭菌天鹅绒转移工具来完成。将转移工具在四环素平板表面轻轻地触一下，沾上每个菌落的细胞，然后以相同的相对位置轻触一个新的氨苄青霉素平板表面，释放细胞，使原初的每个菌落都部分地留在氨苄青霉素平板的相应位置上。然后从四环素平板中挑取那些在氨苄青霉素平板对应位点不能生长的克隆。最后，利用传统有效的科学工具——灭菌牙签，将挑选出的阳性菌落转入新的培养基保存或立即使用。注意我们不把这一过程叫做**选择**（selection），因为没有自动去除不需要的克隆，而是对每个克隆逐个检查，我们把这一更繁琐的过程叫做**筛选**（screen）。

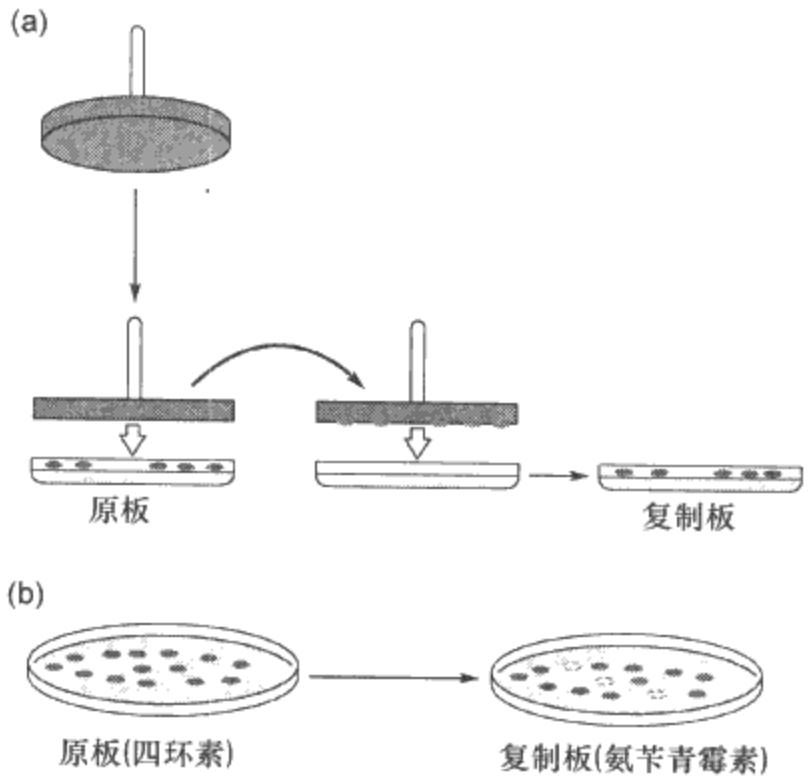
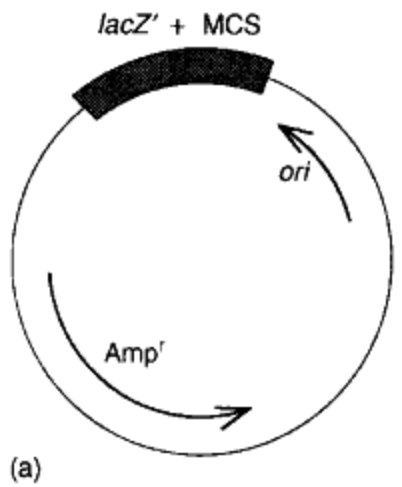
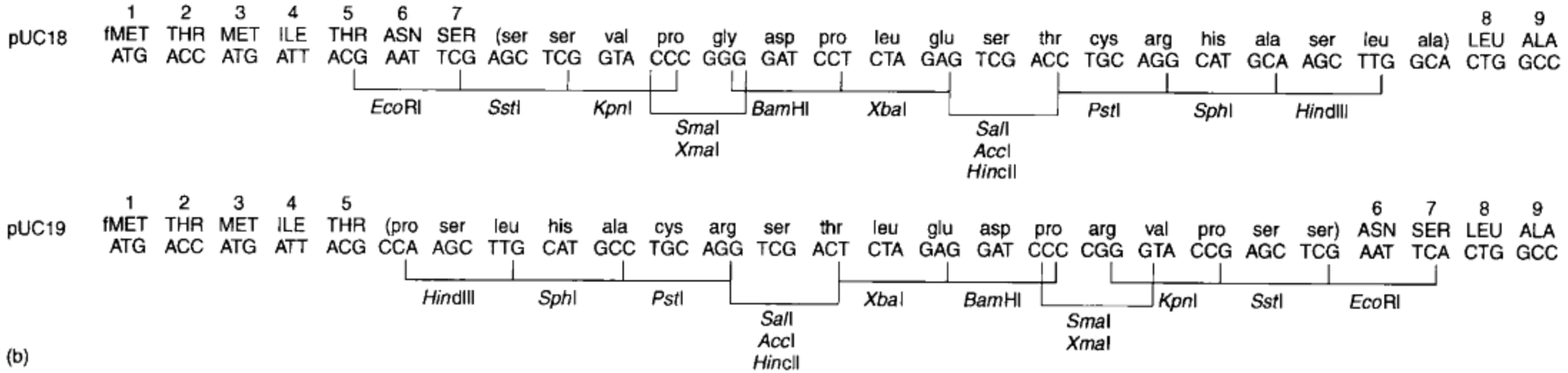


图 4.5　影印培养法筛选细菌。（a）影印法的过程。先用天鹅绒包裹的圆形工具轻轻触一下含有菌落的第一个培养板的表面，每个菌落的细胞粘到天鹅绒上，并可在相同的相对位置上被转移到复制培养板上。（b）影印培养法筛选插入外源基因的 pBR322 氨苄青霉素抗性菌落。所有含有 pBR322 的菌落都在原初的含四环素的平板上生长。复制培养板含氨苄青霉素，凡含有在氨苄青霉素抗性基因中插入外源基因的 pBR322 质粒的菌落则不能在其上生长（表示为空斑）。对应菌落可在原始培养板的相同部位挑出。

现在，已有许多质粒克隆载体可供选择，其中一类虽然过时但很有用的质粒是 pUC 系列（图 4.6）。该系列以 pBR322 为基础，删除了 40% 的 DNA，包括四环素抗性基因。pUC 载体的克隆位点集中在一个小的区域内，叫做**多克隆位点**（multiple cloning site，**MCS**）。pUC 载体含有 pBR322 的氨苄青霉素抗性基因用于筛选接受了载体的细菌。此外，为补偿失去四环素基因，pUC 系列载体含有更便于筛选重组 DNA 的遗传元件。



(a)



(b)

图 4.6　pUC质粒的结构。（a）pUC质粒保留了pBR322的氨苄青霉素抗性基因和复制起点,此外还包含插入了多克隆位点(MCS)的编码β-半乳糖苷酶氨基端的基因(*LacZ′*)。（b）pUC18和pUC19的多克隆位点。pUC18的MCS有13个限制性位点,插入*LacZ′*基因的第7个密码子之后,pUC19的MCS位于第5个密码子之后且是反向的。括号中的小写字母表示由MCS编码。注意,除含有限制性位点之外,MCS保存了*LacZ′*基因的可读框,无外源基因插入时质粒仍具有该基因的活性表达。因此带有单独载体的菌落在含有β-半乳糖苷酶底物X-gal的培养基上会变蓝,而带有载体+插入基因的菌落则保持白色。

图 4.6 给出了 pUC18 和 pUC19 的 MCS，它位于编码 β-半乳糖苷酶氨基端部分（α-肽）的 DNA 序列内（标识为 *lacZ′*），而用于 pUC 载体的宿主菌带有编码 β-半乳糖苷酶羧基部分（ω-肽）的基因。由这些不完整基因产生的 β-半乳糖苷酶片段自身都不具有生物活性，但是它们可以在细菌体内互补，也称为 **α 互补**（α-complementation）。换句话说，这两个不完整基因的产物可以协作形成有活性的 β-半乳糖苷酶。当 pUC18 自己转化带有部分半乳糖苷酶基因的细菌细胞时，可以产生有活性的 β-半乳糖苷酶。如果将这些克隆在含有半乳糖苷酶指示物的平板培养基上培养，那么转有 pUC 质粒的克隆就会改变颜色。指示剂常用 X-gal，它是一种合成的无色半乳糖苷，当 β-半乳糖苷酶切割 X-gal 时，会释放半乳糖和一种蓝色染料，后者可将细菌菌落染成蓝色。

另一方面，如果在多克隆位点插入一段外源基因，中断 β-半乳糖苷酶的这部分基因，使之失活，不能产生与宿主菌的 β-半乳糖苷酶片段互补的产物，则 X-gal 保持无色。因此挑选重组克隆很简单，只需挑选白色菌落即可。注意，与 pBR322 筛选相比，这是个一步筛选过程，我们可同时看见既能在氨苄青霉素培养基上生长又在 X-gal 存在时为白色的克隆。在构建多克隆位点时很小心地保存了 β-半乳糖苷酶基因的可读框，因此即使该基因被 18 个密码子所阻断，仍可产生有功能的蛋白质。但是更大插入片段的进一步阻断通常会破坏 β-半乳糖苷酶基因的功能。

即使利用颜色筛选，插入 pUC 载体的克隆依然会有假阳性发生，即无插入片段的白色菌落。如果在载体与插入片段连接前，载体末端被核酸酶稍微“咬掉”了一点，这些轻微降解的载体在连接这一步自身连接，*lacZ′* 基因就足以产生白色菌落了。因此要使用纯 DNA 和无核酸酶活性的酶。

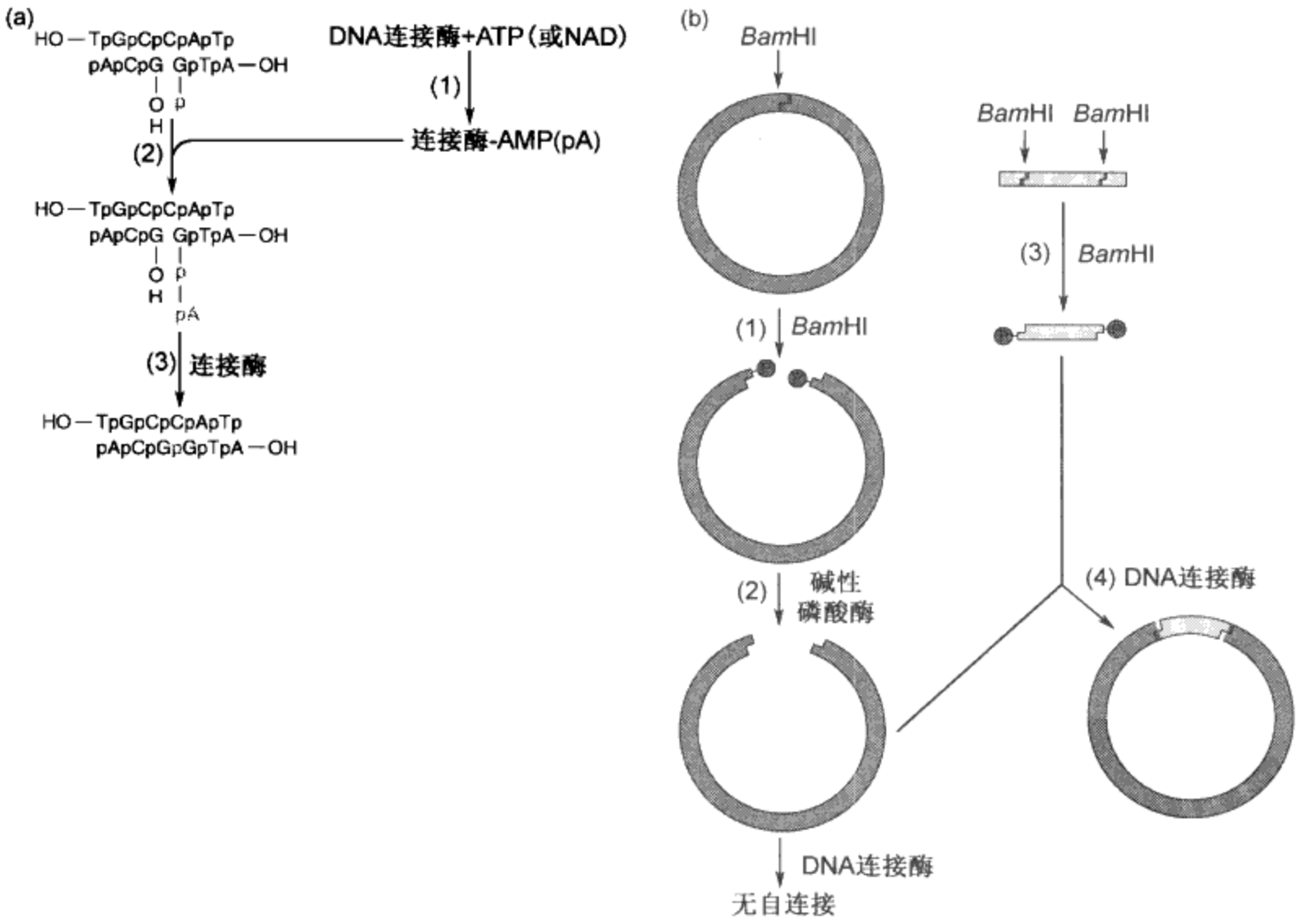


图 4.7　载体与插入片段的连接。（a）DNA 连接酶的机制。步骤（1）：DNA 连接酶和 AMP 供体反应，供体可以是 ATP 或 NAD（烟酰胺腺嘌呤二核苷酸），取决于连接酶的类型。反应生成活性酶（ligase-AMP）。步骤（2）：活性酶将 AMP（pA）供给双链 DNA 的下方链切口处的 5′端自由磷酸基，在切口的一边形成高能二磷酸基。步骤（3）：以磷酸基团间键断裂所产生的能量形成磷酸二酯键封闭 DNA 的切口。这个反应可发生在 DNA 双链上，因此两个独立 DNA 也可被 DNA 连接酶连接起来。（b）碱性磷酸酶防止载体自连接。步骤（1）：用 *Bam*HI 酶切载体，产生带有 5′磷酸的黏性末端。步骤（2）：用碱性磷酸酶去除磷酸基团使载体不能自连接。步骤（3）：也用 *Bam*HI 酶切插入片段，产生带磷酸基的黏性末端。步骤（4）：最后将载体和插入片段连接在一起。插入片段的磷酸基可以形成两个磷酸二酯键（红色）同时留下两个未形成的键或切口，一旦 DNA 转化到细菌细胞这些切口就可被连接。

对于无颜色筛选标记的载体，自连接现象是个大问题，因为更难区分含有插入片段和无插入片段的菌落。即便使用 pUC 或其相关载体，我们也希望减少载体自连接现象。一种比较好的方法是用碱性磷酸酶处理载体，除去自连接所需的 5′-磷酸基团。没有这些磷酸基团，载体就不能自连接，但仍能与含有 5′-磷酸基的外源插入片段连接。图 4.7 显示这一过程。注意，由于只是插入片段有磷酸基，因此在连接产物中还存在两个切口（未形成的磷酸二酯键），但这不成问题，这些切口可以在重组质粒进入细菌细胞后被宿主体内的 DNA 连接酶封闭。

多克隆位点的存在允许用两种不同的酶（如 *Eco*RI 和 *Bam*HI）切割载体，然后将一端为 *Eco*RI 末端，另一端为 *Bam*HI 末端的 DNA 片段克隆到载体上。这种克隆方式叫做**定向克隆**（directional cloning），因为插入到载体上的 DNA 只有一个方向（插入片段的 *Eco*RI 和 *Bam*HI 末端必须与载体上的对应末端相匹配）。知道插入片段的方向有一定的益处，这一点将在本章的后面讨论。定向克隆的另一个好处是防止载体自连接，因为两个限制酶酶切位点的序列无互补性。

注意，pUC18 的多克隆位点刚好与 pUC19 的相反，即相对于 *lacZ′* 基因，限制酶酶切位点的顺序是相反的，表明只需换一下 pUC 质粒就可按任意方向克隆一个片段。现在已经有了更方便的载体，我们将在本章后面部分讨论其中的一些。

小结 pBR322 和 pUC 质粒属于第一代质粒克隆载体。前者有两个抗生素抗性基因以及多个专一性限制酶酶切位点，可用于导入外源基因。大多数酶切位点都中断其中一个抗生素抗性基因，因而可直接用于筛选。用 pUC 质粒进行筛选更简便，该系列质粒含有一个氨苄青霉素抗性基因、一个中断 β-半乳糖苷酶部分基因的多克隆位点。筛选具有氨苄青霉素抗性，又不产生活性 β-半乳糖苷酶使指示剂 X-gal 变蓝的菌落即可。多克隆位点也使外源基因定向克隆到两个不同的限制酶酶切位点间变得很方便。

噬菌体载体 细菌噬菌体是将细菌 DNA 从一个细胞转至另一个细胞的天然载体。因此改造噬菌体使其对所有类型的 DNA 做同样的事就是自然而然的了。与质粒载体相比噬菌体载体有一个天然优点，它对细胞感染的效率要比质粒转化细胞的效率高得多，所以噬菌体载体的克隆产量通常较高。噬菌体载体的重组克隆不是细胞群落而是**噬菌斑**（plaque）（噬菌体在宿主菌平板上清出的一个孔）。噬菌斑由一个噬菌体侵染一个细胞，产生的子代噬菌体从该细胞中裂出，杀死细胞，再感染周围的细胞而形成。这一过程一直持续到肉眼可见的噬菌斑出现。由于形成一个噬菌斑的所有噬菌体都来源于同一噬菌体，所以在遗传上是相同的一个克隆。

λ 噬菌体载体 Fred Blattner 小组通过改造有名的 λ 噬菌体构建了第一个噬菌体载体（第 8 章）。他们移去 λ 噬菌体 DNA 的中间区域，保留其复制必需的基因。移去的噬菌体基因可用外源 DNA 代替。Blattner 以神话中在冥河摆渡的船夫 Charon 的名字将其命名为 **Charon 噬菌体**（Charon phage）。就像 Charon 将灵魂送到阴间一样，Charon 噬菌体携带外源 DNA 进入细菌细胞。船夫 Charon 发音为“Karen”，而噬菌体 Charon 通常发音为“sharon”。对 λ 噬菌体如 Charon 4，更一般的术语是**替换载体**（replacement vector），因为 λ DNA 被移去由外源 DNA 替代了。

与质粒载体相比，λ 噬菌体载体一个明显的优点是可容纳更大的外源 DNA。例如，Charon 4 可接受最大为 20kb 的外源 DNA，这是 λ 噬菌体头部容量的极限。比较而言，传统的质粒载体如果整合这么大的 DNA 片段就无法复制了。什么时候需要这么大的容量呢？λ 噬菌体替换载体一般用于构建**基因组文库**（genomic library），如克隆人类全基因组，无疑需要很多克隆。若每个克隆中的插入片段越大，则所需克隆子的总数就越少。实际上，已经构建了人类和其他物种的基因组文库，λ 替换载体是用于这一目的的常用载体。

除了高容量这一优点外，有些 λ 噬菌体载体还具有对插入片段有最小要求的优点。图 4.8 显示 λ 噬菌体载体具有这一特点的原因：为使 Charon 4 载体可以接受插入片段，用 *Eco*RI 切割 Charon 4，该酶在靠近噬菌体 DNA 中部的 3 个位点进行切割，酶切后产生两个“臂”及两个“填充”片段。之后，产生的左右两臂通过凝胶电泳或超速离心纯化，丢弃填充片段。最后一步是将两臂与插入片段连接，插入片段代替那些丢弃的填充片段。

乍一看，未接收插入片段时两个臂也能连接。这种情况确有发生，但不会形成克隆，因为两臂组成的 DNA 太小，不会被包装到噬菌体中。体外包装噬菌体需要将重组 DNA 与组成噬菌体颗粒所需要的各种物质混合。现在可以购买克隆试剂盒，其中带有纯化的 λ 臂和包装提取物。包装提取物对拟包裹 DNA 的大小有严格要求，除 λ 臂外，DNA 大小必须至少为 12kb 但最大不超过 20kb。

由于每个克隆都含有至少 12kb 的外源 DNA，所构建的文库不应浪费空间去克隆那些没有意义的 DNA。认识到这一点很重要，因为即使每个克隆中含有 12～20kb 的外源 DNA，一个文库也最少需要

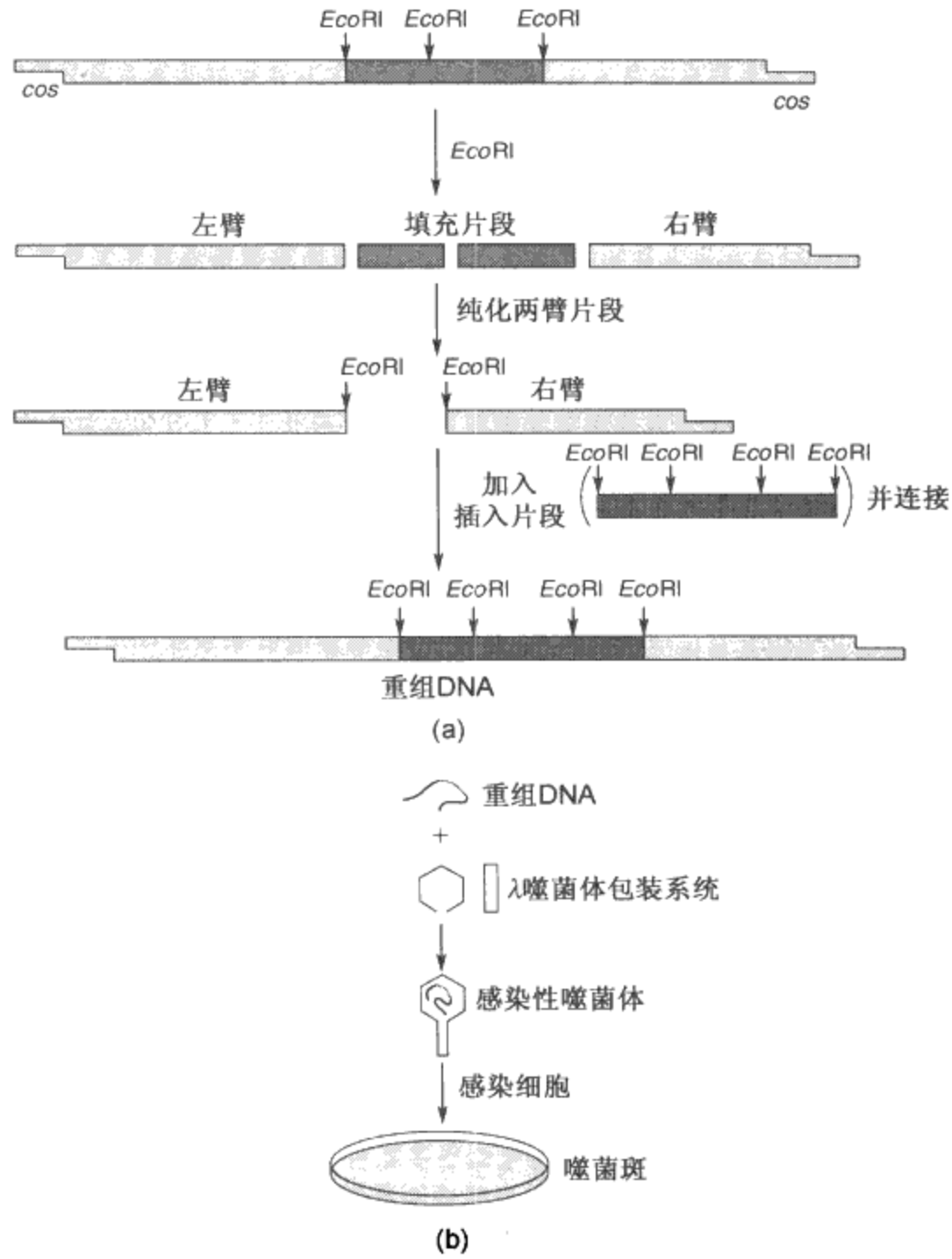


图 4.8　Charon 4 噬菌体克隆。（a）形成重组 DNA。用 *Eco*RI 酶切载体去掉其间的片段留下两臂，然后将不完全消化的插入片段与两臂连接。末端延伸出 12 个碱基的黏端（*cos* 位点），其大小在图上被放大了。（b）包装和克隆重组 DNA。体外将（a）中的重组 DNA 与体外包装提取物混合，其中包括 λ 噬菌体的头部、尾部以及所有将重组 DNA 包装成功能性噬菌体颗粒所需的因子。最后将噬菌体颗粒涂在大肠杆菌平板上，收集产生的噬菌斑。

含有 50 万个克隆才能保证每个人类基因至少出现一次。如果用 pBR322 或 pUC 质粒载体构建人类基因组文库会更加困难，因为细菌会选择性地吸收小的质粒并进行繁殖，所以多数克隆只含有几千个碱基或几百个碱基的插入片段，这样可能就需要数百万个克隆才能完成人类基因组文库的构建。

由于 *Eco*RI 产生平均长度约 4kb 的酶切片段，而 λ 噬菌体载体不接受小于 12kb 的插入片段，因此不能用 *Eco*RI 完全酶切外源 DNA，否则片段会因为太小而不能用于克隆。此外，*Eco*RI 及大部分限制酶能在多数真核基因中部切割一次或多次，所以完全酶切只能得到多数基因的片段。采用 *Eco*RI 不完全酶切可减少这一问题的发生（用低浓度酶或短作用时间，或同时采用这两种方法）。如果 *Eco*RI 每次只在第 4 或第 5 个位点酶切，则产生平均长度16～20kb 的片段，这个大小刚好是 λ 噬菌体载体可接受的，也足以包括大多数真核基因的完整长度。如果想要更随机的片段，可使用机械方法，如超声波破碎法将 DNA 切到适合克隆的大小。

基因组文库一旦构建成功，就可非常方便地研究任何一个感兴趣的基因。唯一的问题是没有目录帮助查找具体的克隆，因此需要利用一定的探针来显示哪个克隆含有目的基因。理想的探针应当是带有标记的核酸分子，其序列与目的基因匹配，然后做一个**噬菌斑杂交**（plaque hybridization）实验，使文库的每个噬菌体 DNA 都与标记的探针杂交，带有标记杂合体 DNA 的噬菌斑就是要挑选的。

在第 2 章我们已经提到杂交反应，在第 5 章将再次讨论。图 4.9 是噬菌斑的杂交原理。在几个培养皿中培养了数以千计的噬菌斑（为简洁起见这里只显示几个）。然后用滤膜与培养皿表面接触［滤膜由 DNA 结合材料如**硝化纤维**（nitrocellulose）或涂层尼龙膜制成］，使每个噬菌斑的部分噬菌体 DNA 转移到滤膜上。然后 DNA 经碱变性后与标记探针杂交，在加入探针之前，将膜用非特异性的 DNA 或蛋白质饱和，其目的是阻止非特异性探针与之结合。当标记探针遇到互补片段（只能是来自目的克隆的 DNA）时与其杂交，标记该 DNA 斑点。然后

用X线胶片检测标记的斑点。X线胶片上的小黑点显示含目的基因的噬菌斑在原始培养皿中的位置。实际情况下，原始平板可能挤满了噬菌斑，不可能准确挑出需要的那个，可从对应位置挑出几个噬菌斑，以很低的噬菌斑密度再涂板一次，然后重复上述杂交过程，找出阳性克隆。

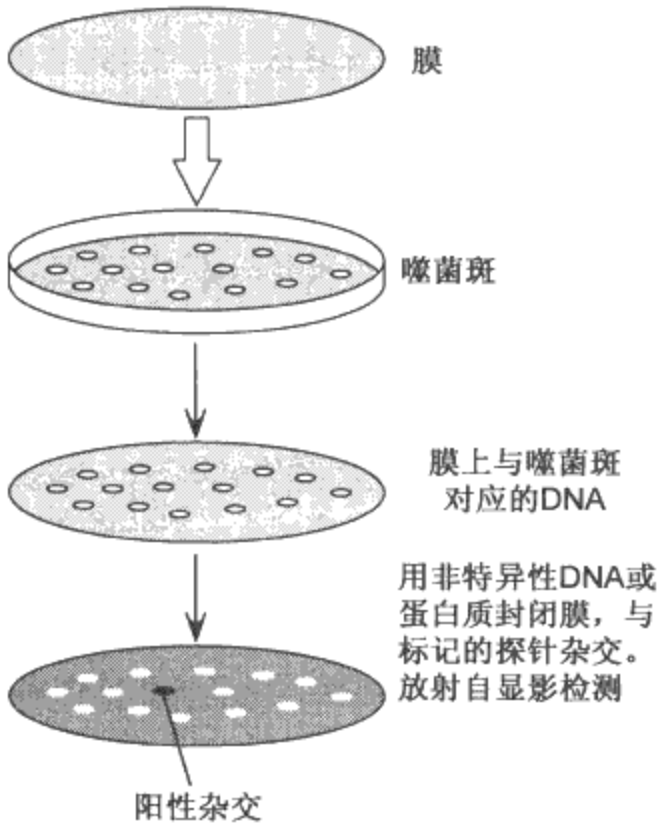


图 4.9　噬菌斑杂交法挑选阳性克隆。用一块尼龙膜或类似介质轻触培养板的表面，其上含有来自图 4.8 的 Charon 4 噬菌斑。从每个噬菌斑上自然释出的噬菌体 DNA 黏附在膜上。然后对 DNA 碱变性，用标记的探针与膜杂交探测目的基因，经 X 线胶片显示标记位置。膜中部的一个噬菌斑被杂交，如胶片上黑点所示。

我们已经介绍了用 λ 噬菌体载体作为基因组克隆的工具，而其他类型的 λ 载体在构建 cDNA 文库中也十分有用，这部分内容将在本章的后面学习。

黏粒　另一种专门用于克隆大片段 DNA 的载体叫**黏粒**（cosmid）。黏粒既有质粒的特点又有噬菌体的特点，它有 λ 噬菌体 DNA 的 ***cos* 位点**（或黏性末端），可将 DNA 包裹进入噬菌体头部（由此得“cosmid”名字的“cos”部分），还带有质粒的复制起点，可以像质粒一样在细菌体内复制（得其名字的“mid”部分）。

由于除了 *cos* 位点以外的几乎所有 λ 噬菌体基因都被剔除了，因此黏粒载体可以容纳更大的插入片段（40～50kb）。一旦插入片段到位，重组黏粒就在体外被包装成噬菌体颗粒。这些颗粒不能以噬菌体形式复制，因为黏粒载体几乎没有噬菌体 DNA，但依然具有感染性，可携带其重组 DNA 进入细菌细胞。一旦进入，由于含有质粒的复制起点，DNA 就能像质粒一样进行复制。

M13 噬菌体载体　还有一类用作克隆载体的噬菌体是丝状噬菌体 M13（长的细丝状）。Joachim Messing 及同事给这种噬菌体的 DNA 加上了与 pUC 载体家族相同的 β-半乳糖苷酶基因片段和多克隆位点。实际上，先构建的是 M13 载体，然后才将有用的多克隆位点直接转到 pUC 质粒上的。

M13 载体的优点主要在于其基因组是单链 DNA，因此克隆的外源片段可以以单链形式回收。我们在本章的后面将看到，单链 DNA 有利于定点突变（site-directed mutagenesis），通过定点突变可以向基因引入特异的、预定的改变。

图 4.10 是如何将一段双链 DNA 克隆到 M13 中

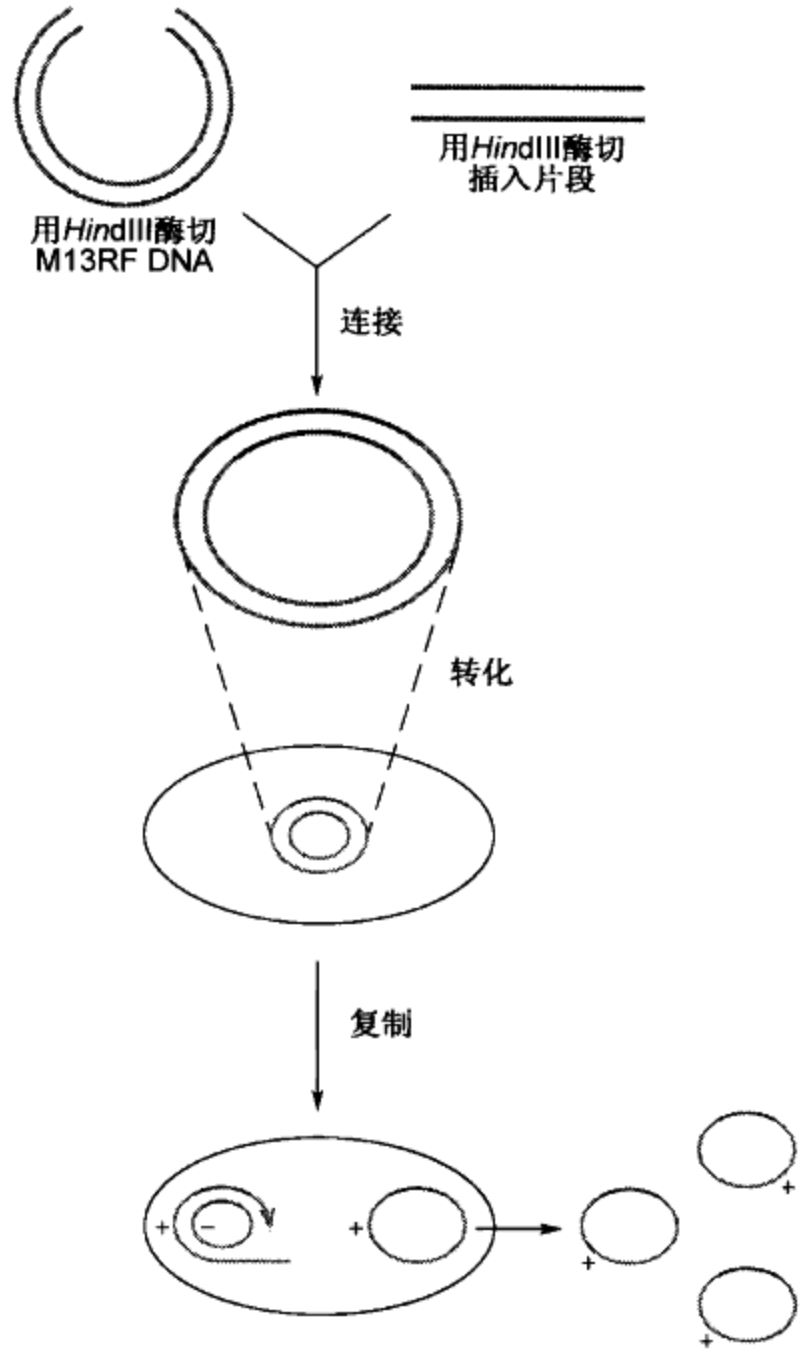


图 4.10　通过 M13 噬菌体克隆获得单链 DNA。用 *Hin*dIII 酶切的外源 DNA 被插入到双链噬菌体 DNA 的 *Hin*dIII 位点。将产生的重组 DNA 转化 *E. coli* 细胞，DNA 在细胞内复制，产生大量单链 DNA 产物。按照惯例 DNA 产物叫做正链（+），模板 DNA 是负链（−）。

获得单链 DNA 的过程。噬菌体颗粒自身的 DNA 是单链的，感染大肠杆菌细胞后，单链 DNA 就转变成双链复制型（replicative form，RF），双链 RF 噬菌体 DNA 可用于克隆。用一个或两个限制酶在多克隆位点酶切后，带有互补末端的外源 DNA 就可插入。然后用这种重组 DNA 转化宿主细胞，产生携带单链重组 DNA 的子代噬菌体。含噬菌体 DNA 的噬菌体颗粒从转化的细胞中分泌出来，并可以从培养基中收集。

噬菌粒　另一种产生单链 DNA 的载体也已开发出来。它们也像黏粒一样既有噬菌体又有质粒的性质，因此被叫做**噬菌粒**（phagemid）。其中常用的一种商品名叫 pBluescript（pBS）（图 4.11）。像 pUC 载体一样，pBluescript 有一个位于 *lacZ′* 基因中的多克隆位点。含插入片段的 pBluescript 可用 X-gal 进行蓝白斑法筛选。同时这个载体还有一个与 M13 相关的单链噬菌体 f1 的复制起点。这意味着含有重组噬菌粒的宿主细胞如果被提供 DNA 复制机器的 f1 辅助噬菌体感染，就会产生并包装单链噬菌粒 DNA。此类载体的另一个优点是多克隆位点的两侧是两个不同噬菌体的 RNA 聚合酶启动子。例如，pBS 的多克隆位点一侧是 T3 启动子，另一侧是 T7 启动子。该特点可用于分离双链重组噬菌粒 DNA，并用任一种噬菌体 RNA 聚合酶进行体外转录产生对应于插入片段的任意一条链的纯 RNA 转录物。

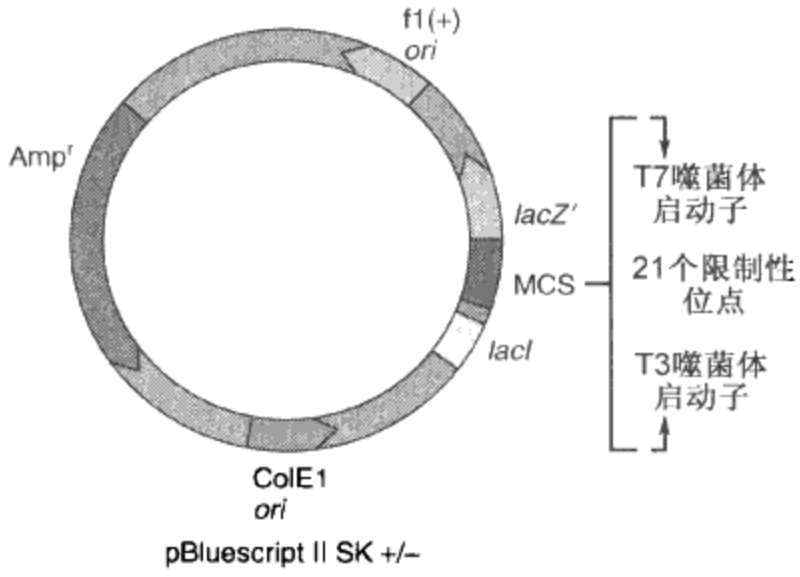


图 4.11　pBluescript 载体。该质粒基于 pBR322 构建，具有其氨苄青霉素抗性基因（Amp^r）和复制起点（*ori*），还具有 f1 噬菌体的复制起点（f1 *ori*）。因此，如果细胞被 f1 辅助噬菌体感染提供了复制机器，载体的单链 DNA 拷贝可包装成子代噬菌体颗粒。多克隆位点（MCS）在两个噬菌体启动子（T7 和 T3）间有 21 个不同的限制酶位点。依据所提供的噬菌体 RNA 聚合酶的种类，DNA 插入片段可在体外转录合成任一条链的 RNA 拷贝。MCS 位于 *E. coli* 的 *lacZ′* 基因中，加入 IPTG 诱导物抵抗 *lacI* 基因表达的阻遏物时，未重组质粒产生 β-半乳糖苷酶的 N 端片段。当加入 X-gal 指示剂时含未重组质粒的菌落显示蓝色。而含有重组质粒的菌落，外源基因在 MCS 中打断了 *lacZ′* 基因，不会产生有功能的 β-半乳糖苷酶，菌落保持白色。

小结　有两类噬菌体载体作为克隆载体非常通用。第一类是 λ 噬菌体载体，从中去除了非必需基因，为插入片段腾出空间。有些工程噬菌体能容纳最大 20kb 的插入片段，因此有利于构建基因组文库，而每个克隆都含有较大片段的基因组 DNA 对基因组文库很重要。黏粒载体能容纳的插入片段更大，可达 50kb，因此也是构建基因组文库的首选载体。第二类噬菌体载体由 M13 噬菌体组成。这类载体有方便的多克隆位点，更有利的是还可产生单链重组 DNA，这一特点可用于 DNA 测序以及定点突变。一类叫做噬菌粒的质粒，也已经被改造成可在其辅助噬菌体存在条件下产生单链 DNA。

真核载体及大容量载体　这类载体可将基因克隆至真核细胞。在本章后面，将介绍几种在真核细胞中表达基因蛋白质产物的载体，还要介绍几种根癌农杆菌（*Agrobacterium tumefaciens*）Ti 质粒设计的载体，这些载体用于将外源基因导入植物细胞。在第 24 章将探讨用于克隆大片段 DNA（数十万个碱基对）的**酵母人工染色体**（yeast artificial chromosome，**YAC**）和**细菌人工染色体**（bacterial artificial chromosome，**BAC**）。

用特异性探针鉴别特定的克隆

从大量克隆中鉴别目的克隆时，探针（特异性地与靶分子结合的分子）是非常必要的。广泛使用的探针有两类：多核苷酸（或寡核苷酸）和抗体。我们先介绍核酸探针，在本章的后面介绍抗体探针。

多聚核苷酸探针　探查目的基因时，可利用已经克隆的其他生物的同源基因。这两个基因的序列要有足够的相似性以便杂交。如果相似性略低，可降低杂交条件的**严谨性**（stringency），以便杂交反应能承受探针与克隆基因间一定程度的碱基错配。

可采取多种方法降低杂交的严谨性。高温、高有机溶剂浓度以及低盐浓度都可促使 DNA 双螺旋解聚。调整这些条件直至只有完全匹配的两条 DNA 链才能形成双螺旋，是高严谨性。放松这些条件（如降低温度）可降低杂交的严谨性，允许有少量错配的两条 DNA 链杂交。

没有同源 DNA 时该用什么呢？如果知道该基因蛋白质产物的至少一段序列时也有办法。我们实

验室在克隆植物毒素蓖麻蛋白（ricin）的基因时就遇到了类似情况，幸运的是蓖麻毒素两条多肽链的全部氨基酸序列都是已知的，因此可以分析氨基酸序列，利用遗传密码简并性推出这些氨基酸可能编码的一系列核酸序列。然后化学合成这些核酸序列并作为探针通过杂交找到蓖麻毒素基因。这种方案是利用由几个核苷酸序列组成的一系列探针，因此称为**寡核苷酸**（oligonucleotide）探针。利用多个寡核苷酸探针是因为遗传密码是简并的，多数氨基酸由一个以上的三联体密码子编码。

幸运的是其中一条蓖麻毒素的多肽包含 Trp-Met-Phe-Lys-Asn-Glu 氨基酸序列，这一特点为我们省去了许多不便。该序列中前 2 个氨基酸各有 1 个密码子，随后的 3 个各有 2 个密码子，第 6 个氨基酸为我们提供了 2 个额外碱基，其简并性只发生在第 3 个碱基上。所以只需要合成 8 种 17 碱基的寡核苷酸（17mer）就可以保证得到这段氨基酸准确的编码序列。这个简并序列可表示如下：

```
              U       G       U
UGG   AUG   UUC   AAA   AAC   GA
Trp     Met     Phe     Lys     Asn   Glu
```

利用这 8 个 17 碱基（UGGAUGUUCAAAAACGA、UGGAUGUUUAAAAACGA 等）的混合物，我们很快找出了几个蓖麻毒素特异性克隆。因此，有了这些准确序列就可以合成出探针。

问题解答

问题

以下是蛋白质的一段氨基酸序列，其基因是你要克隆的

Arg-Leu-Met-Glu-Trp-lie-Cys-Pro-Met-Leu

a. 什么样的五氨基酸序列可给出简并性最小的 17 碱基探针（包括其后密码子的两个碱基）？

b. 要合成多少个不同的 17 碱基才能保证探针与克隆基因的相应序列完全匹配？

c. 如果从最优探针（在 a. 中所选的氨基酸）右移两个密码子处开始探针，那么需要合成多少种不同的 17 碱基？

解答

a. 首先参照遗传密码表（第 18 章）确定上述序列中每种氨基酸的编码简并性，结果如下：

```
6   6   1   2   1   3   2   4   1   6
Arg-Leu-Met-Glu-Trp-Ile-Cys-Pro-Met-Leu
```

每个氨基酸上方的数字表示该氨基酸的简并性。换句话说，精氨酸有 6 个密码子、亮氨酸 6 个、甲硫氨酸 1 个等。现在的任务是找出 5 个连续的简并性最低的氨基酸。我们很快就看出最优组合是 Met-Glu-Trp-Ile-Cys。

b. 为找出你要合成多少种不同的 17 碱基，将氨基酸的简并数相乘即可。对于已选的 5 个氨基酸来说，共有 1×2×1×3×2＝12 种。注意，可以选用 Pro 的前两个碱基（CC）而对简并性没有任何影响，因为编码 Pro 的 4 种简并性均出现在密码子的第三个碱基上（CCU、CCA、CCC、CCG）。因此探针可以有 17 个碱基而不是所选 5 个氨基酸所对应的 15 个碱基。

c. 如果将所选氨基酸向右移两位，也就是以 Trp 开头，那么所需的组合数就是：1×3×2×4×1＝24 种。也就是说共需要制备 24 种探针，而不是 12 种。

小结 对特定克隆可使用与之相互补的寡核苷酸探针进行筛选。知道了某个基因产物的氨基酸序列，可设计一系列编码氨基酸部分序列的寡核苷酸探针。这是筛选具体克隆的最精确高效的一种方法。

cDNA 克隆

cDNA（complementary DNA 或 copy DNA 的简写）是一段 RNA（通常是 mRNA）的 DNA 拷贝。有时我们要构建 **cDNA 文库**（cDNA library），即特定细胞类型在特定时间内全部 mRNA 的一组克隆。这种文库包含数十万个不同的克隆。若想获得一个具体的 cDNA——只含有一个 mRNA 的 DNA 拷贝的克隆，可根据研究目的采取不同的技术。

图 4.12 是一种简单但十分有效的构建 cDNA 文库的方法。任何 cDNA 克隆方法的关键环节都是以 mRNA 为模板利用**反转录酶**（reverse transcriptase）（依赖 RNA 的 DNA 聚合酶）合成 cDNA。反转录酶与所有 DNA 聚合酶类似，没有引物时不能起始 DNA 合成。因此可利用大多数真核生物 mRNA 3′端的 poly（A）尾以 oligo（dT）作为引物。oligo（dT）与 poly（A）互补，以 mRNA 为模板引发 DNA 的合成。

mRNA 被复制产生一条单链 DNA（“第一链”）后，将被**核糖核酸酶 H**（RNase H）部分降解。该酶降解 RNA-DNA 杂交链中的 RNA 链，而我们恰好需要消化掉与第一链 cDNA 碱基配对的 RNA。残留的 RNA 片段作为引物以第一链为模板合成“第二链”。这一阶段要依赖图 4.13 所示的**切口平移**（nick translation）。结果将产生在第二链的 5′端带有小段 RNA 的双链 cDNA。

切口平移就是同时进行切口（单链 DNA 断裂）前 DNA 的移除和切口后 DNA 的合成，就像铺路机一样，铲掉前面老路的同时在后面铺设新路。最终结果是按 5′→3′方向移动或平移切口。用于切口平移的酶是 *E. coli* 的 DNA 聚合酶 I，它具有 5′→3′外切核酸酶活性，随着酶向前移动降解切口前的 DNA。

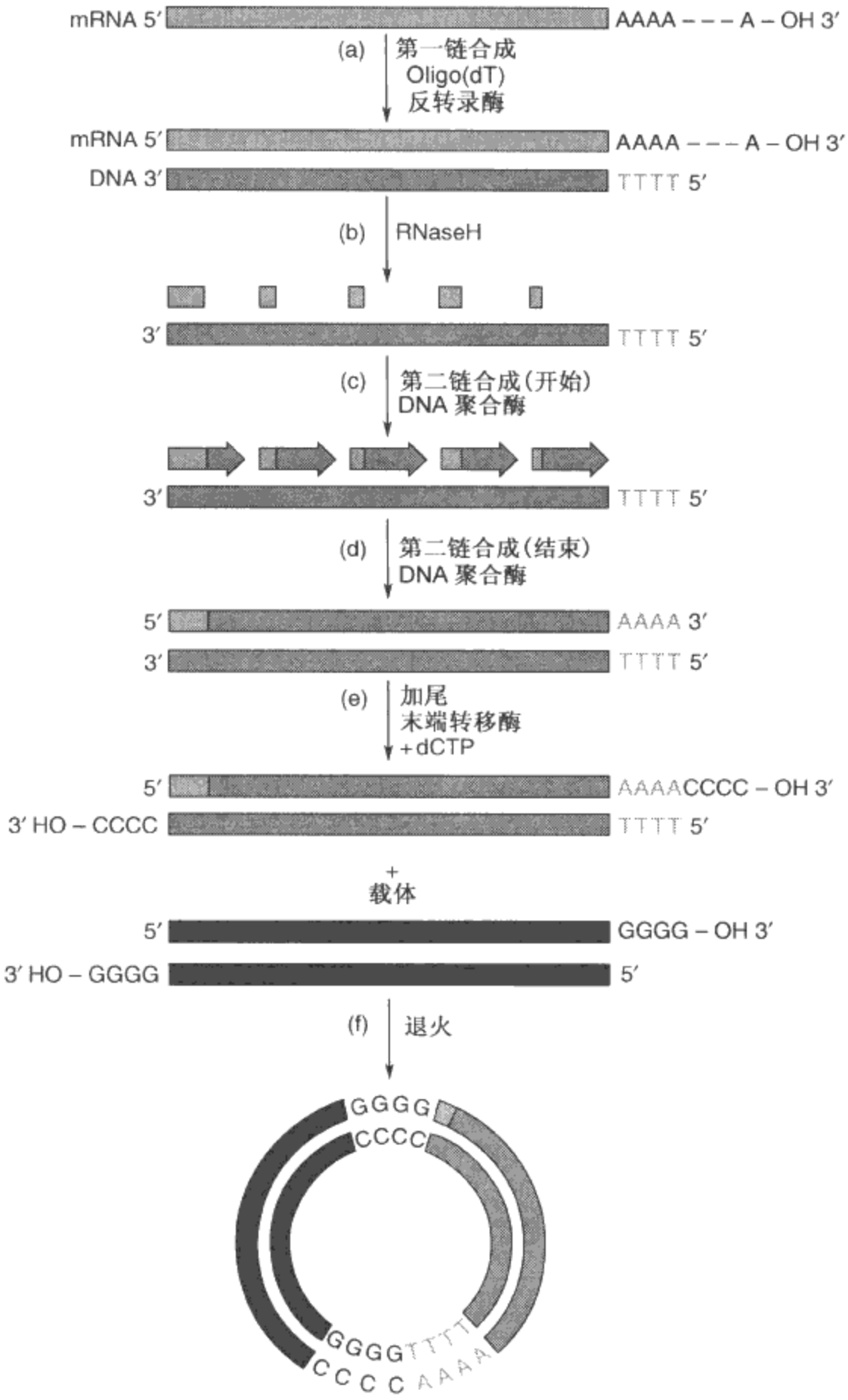


图 4.12　cDNA 文库的构建。重点显示单链 DNA 的克隆，但这种方法也适于 mRNA 混合物构建相应的 cDNA 文库。（a）以 oligo（dT）为引物，反转录酶作用合成的 cDNA 与 mRNA 模板杂交；（b）利用 RNaseH 部分消化 mRNA，形成一系列与 cDNA 第一链碱基配对的 RNA 引物；（c）用 *E. coli* DNA 聚合酶 I 在切口平移条件下依据 RNA 引物合成 cDNA 第二链；（d）cDNA 第二链从最左边的引物开始延长，一路延伸到 3′端的 oligo（dA），以此与 cDNA 第一链的 oligo（dT）引物相对应；（e）给双链 DNA 添加黏性末端，加入末端转移酶和 oligo（dC）；（f）将 cDNA 的 oligo（dC）与适当载体的互补 oligo（dG）末端退火。然后重组 DNA 可以转化细菌细胞，细胞中的酶连接残留的切口并用 DNA 取代剩余的 RNA。

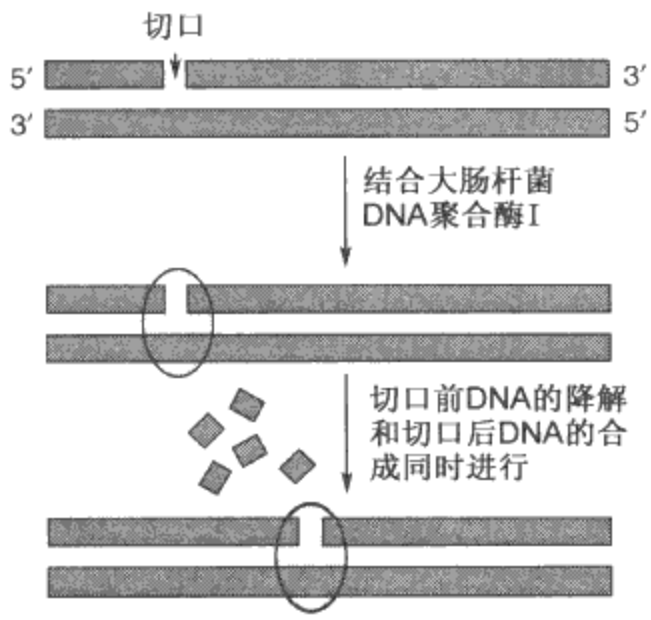


图 4.13　切口平移。本图以普通双链 DNA 为例，但其原理也适用于 RNA 与 DNA 的杂交双链。以双链 DNA 的顶部链有切口开始，*E. coli* DNA 聚合酶 I 结合到切口上，以 5′→3′方向延伸 DNA 片段（从左到右）。同时，其 5′→3′外切核酸酶活性降解右侧的 DNA 片段，为后面产生的片段腾出空位。小方块表示外切核酸酶降解 DNA 后释放的核苷酸。

接下来，将 cDNA 连接到载体上。对于限制酶切割的基因组 DNA 片段来说，这比较容易，但 cDNA不产生黏性末端。虽然效率很低，平末端也可以连接，但最好能获得黏末端的高效连接。用**末端脱氧核苷酰转移酶**（terminal deoxynucleotidyl transferase）[简称**末端转移酶**（terminal transferase）] 和一种脱氧核苷三磷酸可以在 cDNA 上创建黏性末端，本例采用的是 dCTP。转移酶每次将一个 dCMP 加到 cDNA 的 3′端。用同样的方法也可将 oligo（dG）末端加到载体上。将 cDNA 末端的 oligo（dC）与载体末端的 oligo（dG）退火，使 cDNA 和载体连成重组 DNA 可直接用于转化。两个寡核苷酸尾部的碱基配对很强，在转化前不需要连接。被转化细胞的 DNA 连接酶最后会起连接作用，同时 DNA 聚合酶 I 会去除所有残留的 RNA 代之以 DNA。

哪种载体可用于连接 cDNA 呢？有多种选择，取决于检测阳性克隆（带有目的 cDNA 的克隆）的方法。质粒或噬菌粒载体（如 pUC 或 pBS）都可以用，这时阳性克隆要用**菌落杂交**（colony hybridization）鉴定。这个方法与前面介绍过的噬菌斑杂交类似。还可以用 λ 噬菌体，如 λgt11，其 cDNA 置于 *lac* 启动子之下，克隆基因的转录和翻译都可以进行。阳性克隆的鉴定还可利用抗体直接筛选克隆基因的蛋白质产物，在本章的后面将详细讨论这个方法。也可以用寡核苷酸探针与重组噬菌体 DNA 杂交。

cDNA 末端快速扩增

实验中常常得到一些非全长的 cDNA，可能是反转录酶因某种原因未能将 cDNA 扩增至末端。通过**cDNA 末端快速扩增法**（rapid amplification of cDNA end，**RACE**）可将缺失的 cDNA 补齐。图 4.14 是 5′-RACE 技术的原理，用于扩增 cDNA 的 5′端（这是常见的问题），与之相似的 3′-RACE 技术用于填补缺失的 cDNA 3′端。

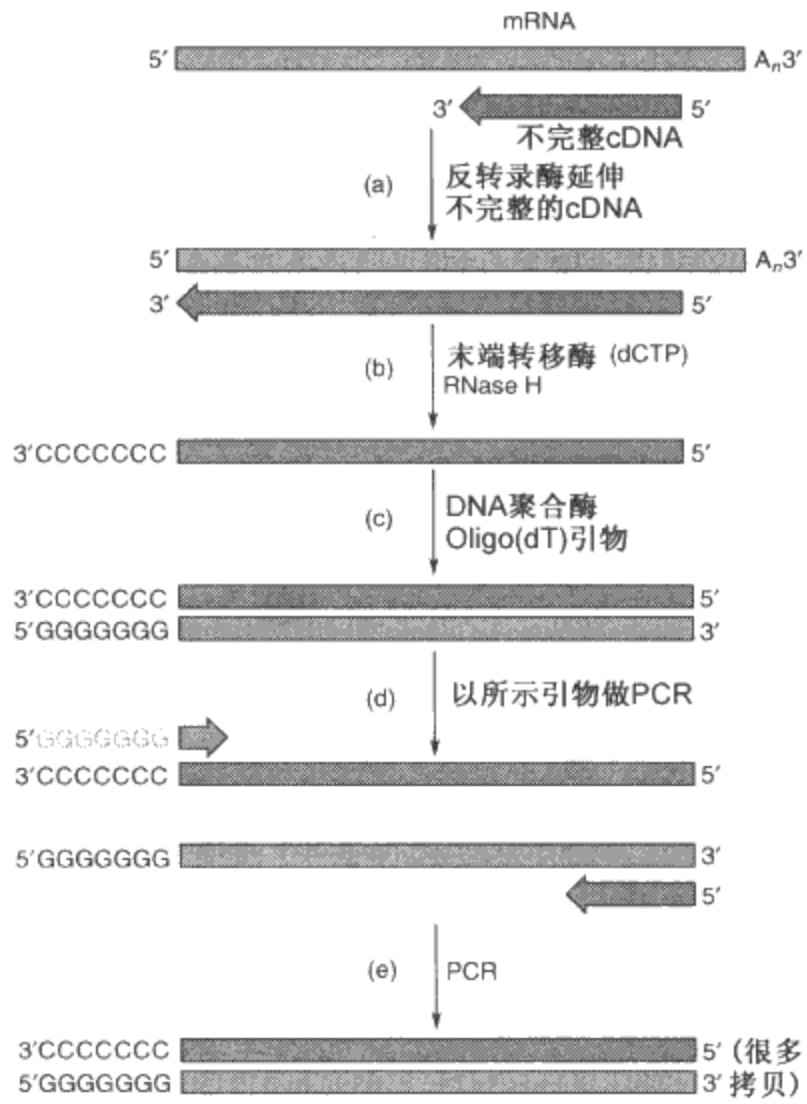


图 4.14　RACE 程序填充 cDNA 的 5′端。（a）将不完整 cDNA 或寡核苷酸片段与 mRNA 杂交，用反转录酶将 cDNA 延伸到 mRNA 的 5′端。（b）用末端转移酶和 dCTP 将 C 添加到延长的 cDNA 的 3′端，再用 RNaseH 降解 mRNA。（c）用oligo（dG）引物和 DNA 聚合酶 I 合成 cDNA第二链。（d）和（e）以 oligo（dG）为正向引物，与 cDNA 的 3′端杂交的寡核苷酸为反向引物进行 PCR。产物就是延伸到 mRNA 5′端的 cDNA。可用类似的程序（3′-RACE）朝 3′方向延伸cDNA，此时无须用末端转移酶给 cDNA 的 3′端加尾，因为 mRNA 已经有 poly（A）了，因此反向引物应该是 oligo（dT）。

进行 5′-RACE 实验时，先制备含有目标序列的 RNA 和 5′端缺失的 cDNA。不完整的 cDNA 链与

mRNA 退火后，用反转录酶复制 mRNA 的其余部分。然后用末端转移酶和 dCTP 给完整的 cDNA 加上 oligo（dC）尾。接下来，用 oligo（dG）引发第二链的合成，该步骤产生一个双链 cDNA。此双链 cDNA 即可用于 PCR 扩增，其引物是 oligo（dG）和 3′端特异性寡核苷酸。

小结 构建 cDNA 文库时，可分先后合成 cDNA的两条链，以某个细胞的 mRNA 为模板合成第一链，然后用第一链为模板合成第二链。反转录酶合成第一链，大肠杆菌 DNA 聚合酶 I 合成第二链。可以给双链 cDNA 加上寡核苷酸尾以便与克隆载体上的尾部碱基互补配对，然后用这些重组 DNA 转化细菌。目的克隆可通过放射性标记的 DNA 探针杂交菌落进行鉴定，或者若使用表达载体（如 λgt11 则以抗体为探针）。不完整的 cDNA 可用 5′-RACE 或 3′-RACE 的方法得到全长。

4.2 聚合酶链式反应

现在我们已经知道如何克隆由限制酶或机械切割所产生的 DNA 片段，也学习了 cDNA 克隆的经典技术。但一种较新的、称为**聚合酶链式反应**（polymerase chain reaction，**PCR**）的技术也可产生用于克隆的 DNA 片段，尤其适用于 cDNA 克隆。

标准 PCR

PCR 技术是由 Kary Mullis 及同事在 20 世纪 80 年代发明的。如图 4.15 所示，该技术使用 DNA 聚合酶合成选定 DNA 区域的拷贝。他们选择扩增 DNA 的 X 片段，通过加入很短的 DNA 为引物与 X 两端的 DNA 序列杂交而引发 X 区域的 DNA 合成。X 两条链的拷贝及最初的 DNA 链又可作为下一轮反应的模板。通过这种方式，所选 DNA 片段的数量每经过一轮都成倍递增，达到初始量的百万倍，直至足够用于电泳检测。

在图 4.15 的例子中，X 的长度是 250bp。注意，第一个循环的产物比 X 要长，因为其合成一直进行到模板 DNA 的末端。只有 3 个循环之后 250bp 的产物才占多数。两个循环之后只有 2 条单链的 250bp 产物，标识在图的底部，一条蓝色线一条红色线。

起初，由于标准 DNA 聚合酶不耐高温（超过 90℃）而每一轮复制之前都需要高温打开 DNA 双链，工作人员必须在每次循环后加入新鲜 DNA 聚合酶。现在有了专门的热稳定性 DNA 聚合酶，其中一种叫 ***Taq* 聚合酶**（*Taq* polymerase），来源于热喷泉中的嗜热杆菌（*Thermus aquaticus*）。实验时只需将 *Taq* 聚合酶与引物和 DNA 模板在试管中混合，密封试管，然后放入**热循环仪**（thermal cycler）中即可。热循环仪被设定在 3 个不同温度反复循环：首先是高温（约 95℃）使 DNA 解链，然后是相对较低的温度（约 50℃）使引物与模板 DNA 退火，最后是中等温度（约 72℃）使 DNA 合成。每个循环只需几分钟，通常不到 20 个循环就可产生大量所需扩增的 DNA。PCR 是如此强有力的扩增体系，甚至还帮助创作了多个科幻小说故事，如《侏罗纪公园》（*Jurassic Park*）（见知识窗 4.1）。

小结 PCR 用于扩增两个预设位点之间的 DNA 片段，与这两个位点互补的寡核苷酸可作为两位点间 DNA 拷贝合成的引物。每一轮 PCR 都使被扩增 DNA 的拷贝数加倍，直到合成大量的片段。

cDNA 克隆中反转录 PCR（RT-PCR）的应用

如果想克隆一段已知 mRNA 序列的 cDNA，可使用**反转录 PCR** 方法（reverse transcriptase PCR，**RT-PCR**），如图 4.16 所示。RT-PCR 与 PCR 技术的主要区别在于它从 mRNA 而不是双链 DNA 开始。因此，需先将 mRNA 反转录成 DNA。这一步由反转录酶和反向引物完成，将 mRNA 反转录成单链 DNA，然后用正向引物将单链 DNA 转换成双链的，然后用常规 PCR 技术扩增 cDNA 用于以后的克隆。扩增 cDNA 时，可设计引物含有限制性位点以便在 cDNA 末端添加限制酶酶切位点。本例中，其中一个引物带有 *Bam* HI 位点，另一个带有 *Hin*dIII位点（在末端添加几个保护性核苷酸可使限制酶有效地切割）。因此 PCR 产物是两个末端都带有限制酶切位点的 cDNA。用这两个限制酶切割 PCR 产物产生黏性末端与所选载体连接。如果设计两个不同的黏性末端就可进行定向克隆，cDNA 在载体上只能是两种可能方向中的一个。这对于将 cDNA 克隆到表达载体非常有用，因为 cDNA 必须与驱动其转录的启动子方向一致。有必要提醒，cDNA 本身不能包含其末端所添加的限制酶酶切位点。否则，限制酶在 cDNA 内部和末端都进行切割，产物将毫无用处。

小结 RT-PCR 可用于以单链 mRNA 为模板合成 cDNA，此时需已知 mRNA 序列以便设计引物。可在 PCR 引物上设置限制酶酶切位点，cDNA的末端含有这些酶切位点。这样易于酶切 cDNA 并将其连接到载体上。

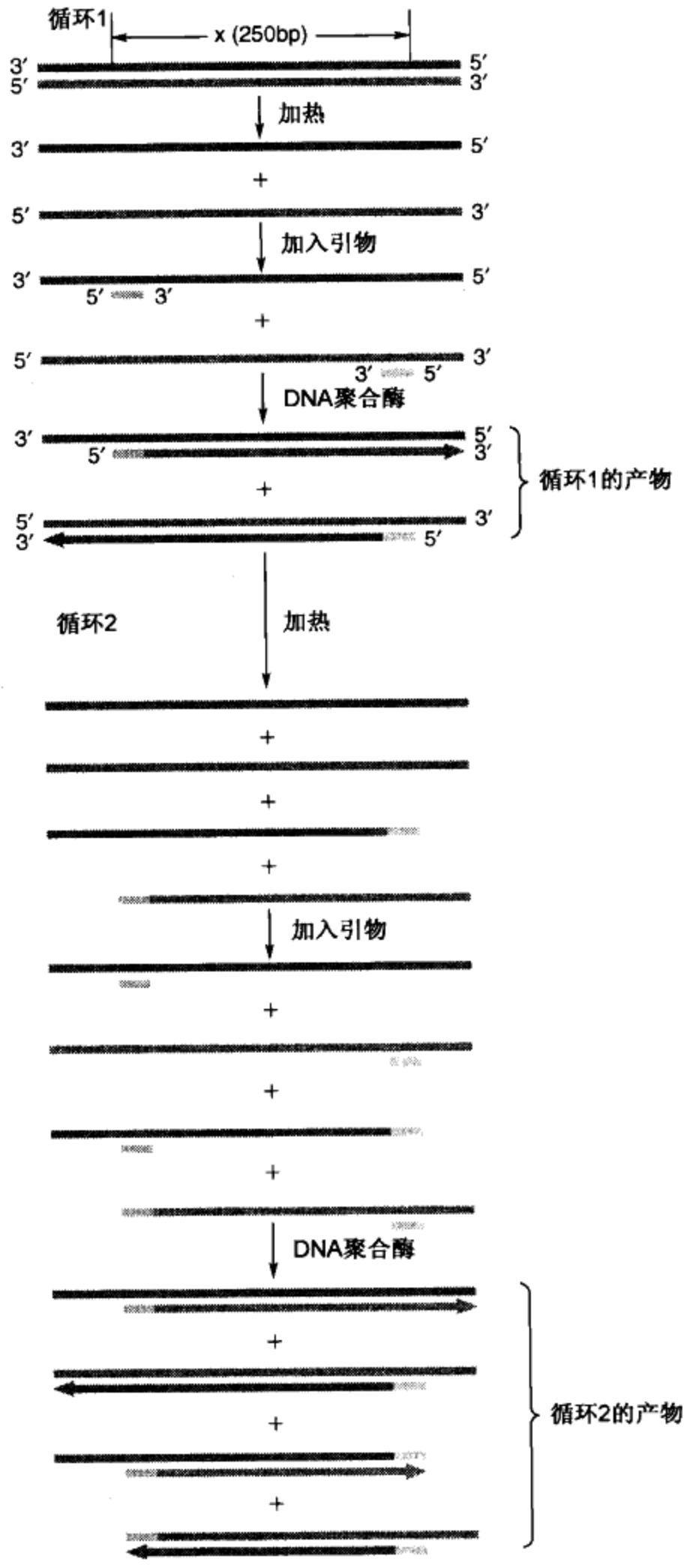


图 4.15　通过聚合酶链式反应扩增 DNA。循环 1：以 DNA 双链开始，加热使两条链分开。然后添加与待扩增区域（X，250bp）两侧互补的单链 DNA 引物。引物分别与单链 DNA 的适当位点杂交，这时热稳定的 DNA 聚合酶用这些引物合成互补的 DNA 链。箭头代表新合成的 DNA，复制在箭头尖处停止。循环 1 结束时呈现 2 个双链 DNA，包括要扩增的区域，而开始时只有一个双链 DNA。循环 1 显示所有 DNA 链和引物都是 5′→3′ 方向。同样原理也适用循环 2。循环 2：重复这个过程，加热分离 DNA 双链，降温与引物退火使热稳定的 DNA 聚合酶合成更多的 DNA。现在 4 条单链 DNA（包括 2 条新合成的）都可作为模板合成互补链。结果得到的 4 个双链 DNA 都含有扩增区域。注意：每轮循环能使 DNA 分子数量加倍，因为每次循环产生的双链 DNA 加入两条母链又作为下一轮复制的模板。这种指数式增长在下一个循环产生 8 个分子，再下轮可产生 16 个。这个过程显然在很短时间内就可产生很大的数量。

知识窗 4.1

侏罗纪公园：不只是幻想?

在 Michael Crichton 的《侏罗纪公园》一书及同名电影中，描述了一位科学家和一位企业家合作制造出活恐龙的神奇故事。他们的策略是分离恐龙 DNA，但不是直接从恐龙残骸中分离。他们发现侏罗纪时期的吸血昆虫吸了恐龙血之后又被树的分泌液包埋变成琥珀，从而保存了这些昆虫。他们认为，既然血液中含有 DNA 的白细胞，昆虫肠道的内含物中就应该含有被吸过血的恐龙 DNA。下一步是用 PCR 扩增恐龙 DNA，将得到的 DNA 片段连接在一起，然后将这些 DNA 置于卵细胞中，瞧啊！一个恐龙就孵化出来了。

这个设想听起来很神奇，但实际上某些现实问题使这种设想还完全停留在科幻领域。令人诧异的是这个故事中的某些部分已经出现在了科学文献中。1993 年 6 月，《侏罗纪公园》在电影院上映的同一个月，《自然》杂志的一篇文章介绍了对一种灭绝象鼻虫基因进行 PCR 扩增和部分基因测序的结果，这种昆虫在黎巴嫩的琥珀已经包埋了 1.2 亿～1.35 亿年。这一研究把我们带回了白垩纪，虽然没有侏罗纪那么远古，但白垩纪也是恐龙的世界。如果这种研究可行，也许真能找到肠道中含有恐龙 DNA 的保存下来的吸血昆虫呢，进一步说，得到的 DNA 有可能仍然足够完整，可作为 PCR 扩增的模板。毕竟，PCR 技术有能力将单一 DNA 分子扩增到我们所期望的任何程度。

那么制造恐龙的道路上还有什么障碍呢？先不考虑用裸 DNA 制造脊椎动物这一未知领域，我们首先得考虑 PCR 技术自身的限制，其中之一是 PCR 扩增 DNA 片段的大小是有限的，至多 40kb。这样的长度只是整个恐龙基因组的成千上万分之一，这就意味着我们得将至少成千上万的 PCR 片段一段一段地连接起来重建整个基因组，而完成这一任务的前提是我们要对恐龙 DNA 的序列有足够的认识，才能在开始时为所有这些片段设计 PCR 引物。

但是如果我们能设法让 PCR 扩增片段远不止 40 000bp 会怎么样？如果 PCR 可以扩增整个染色体，即一次扩增高达几亿个碱基对呢？即便如此，我们还会面临 DNA 是一种遗传上不稳定的分子这一事实，就算是在琥珀里的昆虫体内也不能指望全长染色体能保存几百万年。PCR 之所以能扩增相对较短的片段是因为引物只需要找到一个在这一短长度上完整的分子就可以了，而要找到一个完整无断裂的染色体，哪怕是一段无断裂的百万碱基 DNA 片段，都是不可能的。

这些因素给少数已公布的用 PCR 扩增远古时代 DNA 的例子增添了相当的不确定性。很多科学家认为像 DNA 这种容易断裂的分子能够保存数百万年根本不可信。他们认为恐龙 DNA 很久以前就分解为核苷酸了，根本不能用作 PCR 扩增的模板。实际上，所有超过 10 万年的远古 DNA 都是一样，包括原来认为可能是例外的保存在琥珀中的 DNA。

另外，PCR 仪确实扩增了远古昆虫样品的某种 DNA。如果这不是远古昆虫的 DNA，又是什么呢？这涉及到 PCR 方法的第二个限制，也是其最大优点：它的极高灵敏度。正如我们所知，PCR 能扩增单一 DNA 分子，如果这是我们想要扩增的 DNA 就再好不过了。然而，如果样品中只要含有一个分子的混杂 DNA，PCR 就能对其扩增而产生污染。

有鉴于此，检测白垩纪象鼻虫 DNA 的研究者先完成对白垩纪象鼻虫 DNA 的所有 PCR 扩增和测序，然后再利用现代昆虫 DNA 来与白垩纪象鼻虫 DNA 进行比较。这样才可避免在扩增所谓灭绝象鼻虫基因时因以前实验留下的现代昆虫 DNA 的污染。但是 DNA 无处不在，尤其是在分子生物学实验室，想消除每一个遗留的 DNA 分子是极困难的。

此外，在昆虫肠道内的恐龙 DNA 会被该昆虫的 DNA 严重污染，更不用说来自肠道细菌的 DNA 了。而且谁敢说这个昆虫在死于树脂包埋之前只吸了一种恐龙的血呢？如果它吸了两种恐龙的血，PCR 会将这两种恐龙的 DNA 一起扩增，我们也根本无法将这两种 DNA 分开。

换句话说，就如想象的能看到一个活生生恐龙那么让人兴奋一样，我们已经拥有了创造一个真实侏罗纪公园的工具。但是，现实问题使这一理想几乎不可能。

在更为现实的水平上看，PCR 技术已经使我们可以对灭绝生物与其现存近亲的基因序列进行比较。由此催生了一个令人振奋的新领域，植物学家 Michael Clegg 称之为“分子古生物学”。

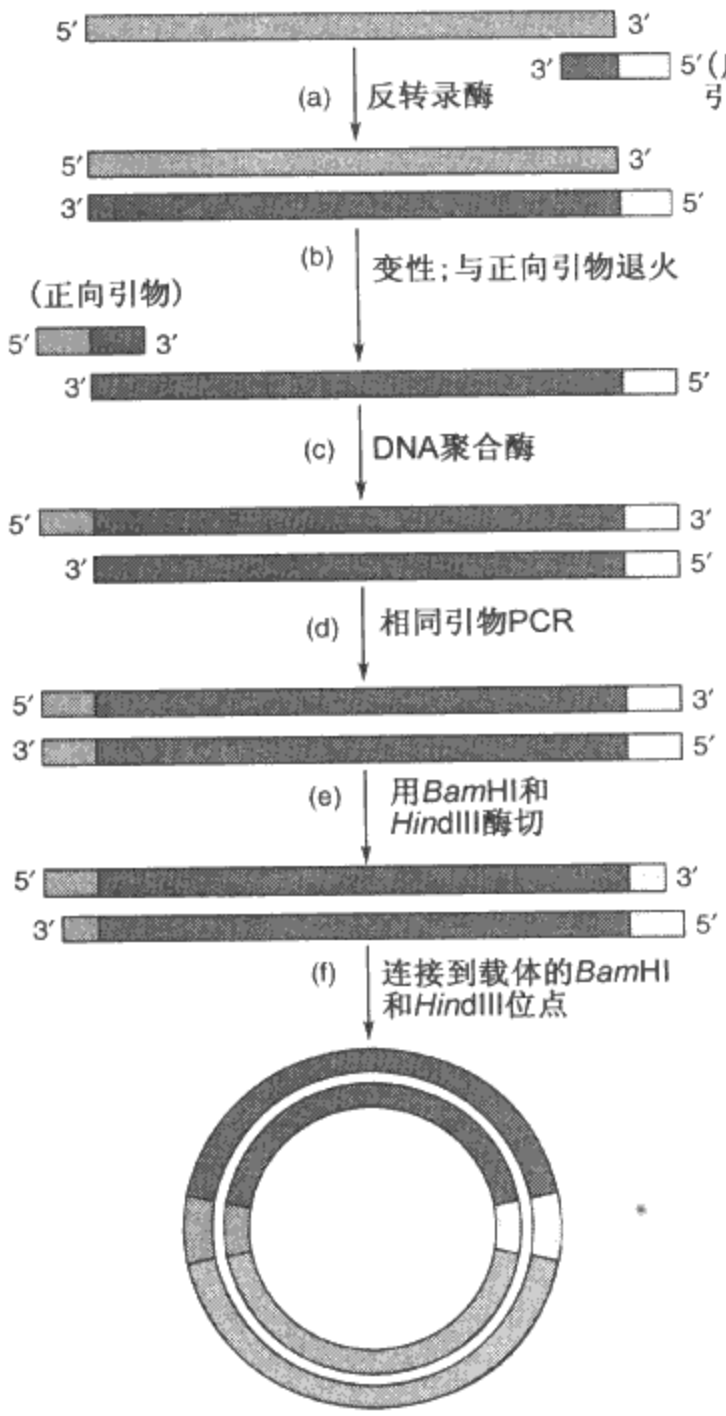


图 4.16　利用 RT-PCR 克隆单链 cDNA。（a）5′端带有 *Hin*dIII 位点的反向引物在反转录酶催化下合成 cDNA 第一链。（b）将 DNA-RNA 杂交链变性后与 5′端带有 *Bam*HI 位点的正向引物退火。（c）在 DNA 聚合酶 I 催化下，正向引物起始 cDNA 第二链合成。（d）用相同的引物继续 PCR 扩增双链 cDNA。（e）用 *Bam*HI 和 *Hin*dIII 酶切 cDNA 产生黏性末端。（f）把 cDNA 连接到适当载体的 *Bam*HI 和 *Hin*dIII 位点上。最后用重组 cDNA 转化细胞获得克隆。

实时定量 PCR

实时定量 PCR（real-time PCR）是一种在 DNA 扩增时（即实时）对其定量的方法。图 4.17 是实时定量 PCR 的原理。两个 DNA 链分离后，不仅与两端的引物还与**报告探针**（report probe）一起退火。报告探针是荧光标记的、与 DNA 链的一部分互补的寡核苷酸。其 5′端有荧光标签（F），3′端有荧光猝灭标签（Q）。

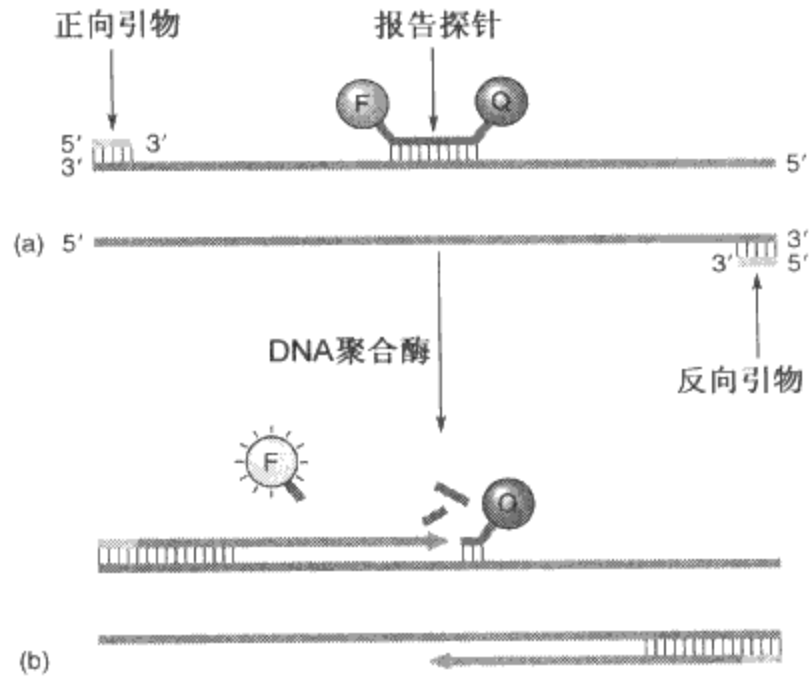


图 4.17　实时定量 PCR。（a）正向和反向引物与两条分开的 DNA 链退火，报告探针与顶链退火。报告探针的 5′端有一个荧光标签（F），3′端有一个猝灭标签（Q）。（b）DNA 聚合酶已将引物延长（箭头表示），同时为给复制顶链开路还降解了部分报告探针，使荧光标签与猝灭标签分离而发出其正常荧光。复制的 DNA 链越多，观察到的荧光也越多。

在 PCR 延伸这一步，DNA 聚合酶延伸正向引物，遇到报告探针后开始对其降解，以便在这段区域合成新的 DNA。随着报告探针的降解，荧光探针与猝灭探针分离，荧光强度迅速增加。整个过程在荧光仪中进行，可以检测标签的荧光强度，以此作为对 PCR 反应过程的度量。有足够的报告探针与新合成的 DNA 链退火，因此荧光强度随着每一轮扩增而增加。

遗憾的是“real-time”和“reverse transcriptase”都简写为“RT”，因此在文献中看到“RT-PCR”时需要看上下文搞清楚是哪一种 PCR。也可以做实时反转录 PCR，从 RNA 而不是双链 DNA 开始，这种方法可以简写为**“实时反转录 PCR”**（real-time RT-PCR）。

小结　实时定量 PCR 通过检测与正向引物互补链杂交的报告探针的降解而跟踪 PCR 过程。探针降解时荧光标签与猝灭标签分离，荧光强度增加，其强度可在荧光仪中实时检测。

4.3　表达克隆基因的方法

为什么要克隆基因？本章开始讲到，因为基因克隆可以使我们获得大量特定 DNA 序列以便进行

详细研究。因此，基因本身就是基因克隆的有价值产物。基因克隆的另一个目的是生产大量的基因产物，既可用于研究，也可创造利润。

如果目的是利用细菌生产所克隆真核基因尤其是高等真核生物基因的蛋白质产物，利用 cDNA 可能比直接从基因组中切出的基因效果要好。这是因为大多数高等真核生物的基因都含有叫做内含子（第 14 章）的间隔，而细菌对此无法处理。真核细胞通常能转录这些间隔形成前体 mRNA，然后把它们剪掉，并把剩余的部分（即外显子）拼接起来形成成熟 mRNA。因此作为 mRNA 拷贝的 cDNA 已经剔除了内含子，可以在细菌细胞中正确表达。

表达载体

到现在为止，我们所讨论的载体主要用于克隆的第一阶段，先将外源 DNA 导入细菌并使其复制。大多数情况下，这一过程很有效，克隆载体在 *E. coli*中稳定生长并产生大量的重组 DNA。有些克隆载体还可作为**表达载体**（expression vector）生产克隆基因的蛋白质产物。例如，pUC 和 pBS 载体将插入的 DNA 置于多克隆位点上游的 *lac* 启动子控制之下。如果插入的 DNA 恰好与其打断的 *lacZ′* 基因在同一个可读框内，那么将产生一个**融合蛋白**（fusion protein）。其氨基端是 β-半乳糖苷酶的部分蛋白质序列，羧基端则是插入 DNA 所编码的蛋白质序列（图 4.18）。

如果对所克隆基因的高效表达有兴趣，则专用表达载体的效果通常会更好。典型的细菌表达载体含有基因有效表达所必需的两个元件：一个强启动子和一个靠近起始密码子 ATG 的核糖体结合位点。

可诱导型表达载体　表达载体的主要功能是产生基因产物，产物越多越好。因此，表达载体通常都配以很强的启动子，其理论基础是产生的mRNA越多，其蛋白质产物也越多。

使克隆的基因保持抑制状态，当需要的时候再表达会很有益处。原因之一是在细菌中产生大量真核蛋白会有毒害作用。即使这些蛋白质并不是真正有毒的，但是积累到巨大量就会干扰细菌的生长。不管何种情况，如果所克隆的基因持续处于开启状态，那么含有该基因的细菌不会长到足够浓度来生产有实际意义的蛋白质产量。在细菌中过高表达出现的另一问题是表达的蛋白质会形成叫做**包含体**（inclusion body）的不溶性聚合体。因此，将所克隆的基因置于可诱导启动子的下游有助于使其处于关闭状态。

lac 启动子在一定程度上是可诱导的，期望在被人工诱导物 IPTG 激活前是关闭的。但 *lac* 阻遏物的

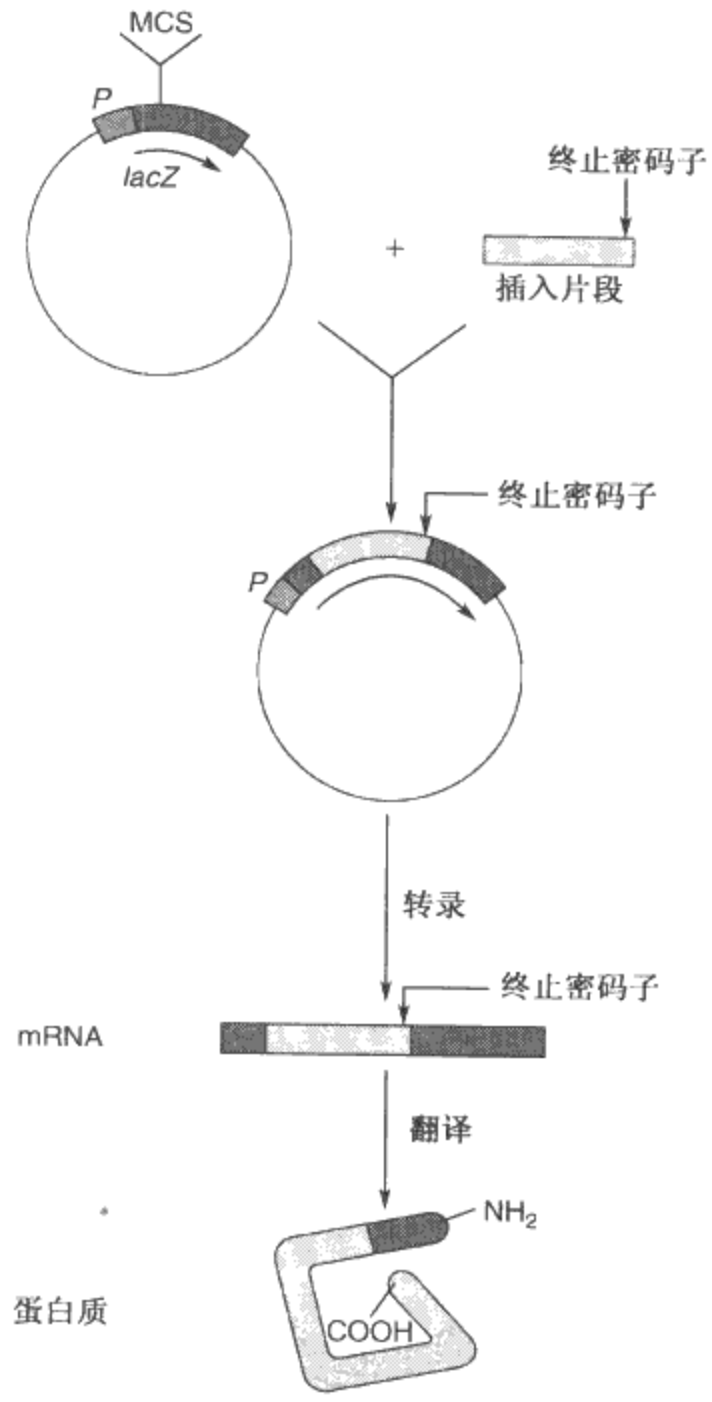


图 4.18　通过 pUC 质粒克隆获得融合蛋白。将外源 DNA 插入 MCS，从 *lac* 启动子（P）处开始的转录产生了杂合的 mRNA：从几个 *LacZ′*密码子开始，变到插入序列，然后又回到 *LacZ′*。这个 mRNA 被转录成融合蛋白：开始（氨基端）是几个 β-半乳糖苷酶的氨基酸，接着是插入的氨基酸序列组成了融合蛋白。因为插入序列含有翻译终止密码子，所以其余的 *LacZ′*密码子不翻译。

抑制作用不完全（有漏缝），在无诱导物条件下也能检测到克隆基因的表达产物。一种解决办法是在自身携带 *lacI*（阻遏物）基因的质粒或噬菌粒中表达基因，如 pBS（图 4.11）。载体产生的阻遏物使所克隆的基因被关闭直到被 IPTG 诱导（关于 *lac* 启动子见第 7 章）。

但是 *lac* 启动子并不是太强，因此许多载体都设计带有一个杂合的 ***trc* 启动子**（*trc* promoter），该启动子结合了 *trp*（色氨酸操纵子）启动子的强启动能力和 *lac* 启动子的可诱导性。由于－35 框的存在，

trp 启动子比 *lac* 启动子强得多（第 6 章）。于是，分子生物学家将 *trp* 启动子的－35 框与 *lac* 启动子的－10 框结合起来，同时加上 *lac* 操纵基因（第 7 章）。*trp* 启动子的－35 框使杂合启动子很强，*lac* 操纵基因使其可被 IPTG 诱导。

阿拉伯糖操纵子（*ara*）的启动子 P_{BAD}，可对转录精确调控。该启动子受阿拉伯糖诱导（第 7 章），因此没有阿拉伯糖时不发生转录，当添加到培养基中的阿拉伯糖越来越多时，转录就会越来越强。图 4.19 以实验说明了这个现象。把绿色荧光蛋白基因克隆到一个 P_{BAD} 载体中，其表达随着阿拉伯糖浓度的升高而被诱导。缺乏阿拉伯糖时，无绿色荧光蛋白出现，阿拉伯糖浓度从 0.0004%开始，蛋白质产量逐步升高。

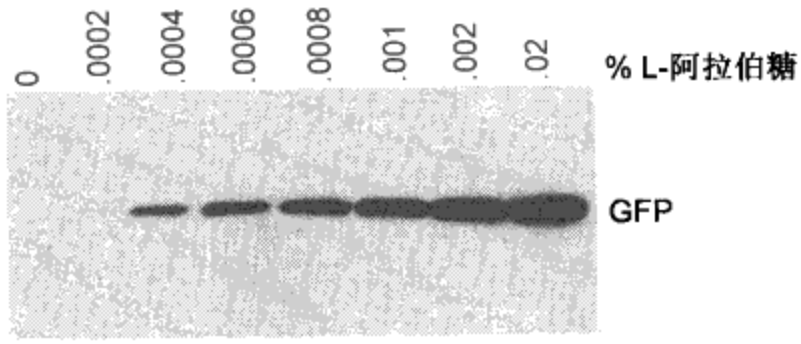


图 4.19　P_{BAD} 载体的使用。将绿色荧光蛋白（GFP）基因克隆到载体的 P_{BAD} 启动子下游。启动子活性通过增加阿拉伯糖的浓度而被诱导。对不同浓度（图上方所示）阿拉伯糖诱导细胞的提取物做电泳，监测 GFP 产物。然后将蛋白质产物印迹到膜上，用抗-GFP 抗体检测 GFP（免疫印迹，第 5 章）。（*Source*：Copyright 2003 Invitrogen Corporation. All Rights Reserved. Used with permission.）

另一种策略是使用严谨调控的启动子，如 λ 噬菌体启动子 P_L。带有该启动子-操纵基因系统的表达载体被克隆到含有温度敏感型 λ 阻遏物基因（*cI857*）的宿主细胞中。只要这些细胞保持在相对较低的温度下（32℃），阻遏物就发挥功能，基因不表达。当温度升至非允许水平（42℃）时，温度敏感型阻遏物不再起作用，对克隆的基因抑制被解除。

为确保目的基因受到严谨调控和高诱导表达，常用的方法是将其置于含 T7 噬菌体启动子的质粒中。然后将质粒导入含有对 T7 RNA 聚合酶基因严谨调控的细胞中。例如，T7 RNA 聚合酶基因在细胞中受修饰过的 *lac* 启动子调控，而该细胞也携带 *lac* 阻遏物基因。因此，除非存在 *lac* 诱导物，否则 T7 RNA 聚合酶基因会受到强烈抑制。如果没有 T7 RNA 聚合酶，目的基因就不能转录，因为 T7 启动子绝对需要它自己的聚合酶。一旦加入 *lac* 诱导物，细胞就开始合成 T7 聚合酶，目的基因开始转录。由于产生了大量 T7 聚合酶，目的基因会被高水平地转录，产生出大量的蛋白质产物。

小结　设计表达载体用于产生所克隆基因的蛋白质产物，一般是尽可能地最大量表达。为此，表达载体包含强的细菌或噬菌体启动子、细菌核糖体结合位点，该位点在所克隆的真核基因中是缺失的。大多数克隆载体（译者注：应为表达载体）都是可诱导的，由此可避免因外源产物提前过量表达而对细菌宿主细胞产生的毒害作用。

产生融合蛋白的表达载体　大多数表达载体都是产生融合蛋白。乍一看这似乎不利，因为没有产生导入基因的天然产物。然而，融合蛋白上的额外氨基酸在纯化蛋白质产物方面是非常有用的。

我们来看一下商品名为 pTrcHis 的寡聚组氨酸表达载体（图 4.20）。这种载体的多克隆位点上游有一个短序列，编码 6 个串联的组氨酸，在这种载体上所表达的蛋白质都是带有 6 个组氨酸的融合蛋白。为什么要把 6 个组氨酸加到蛋白质上呢？这是因为寡聚组氨酸区对镍（Ni^{2-}）等二价金属离子具有很高的亲和性，因此带有这段区域的蛋白质可以利用镍**亲和柱层析法**（affinity chromatography）进行纯化。该方法的独到之处在于其简便和快速。当细菌产生融合蛋白后，只需裂解菌体，把细菌粗提液加到镍亲和柱上，洗脱未结合到柱上的蛋白质，然后用组氨酸或组氨酸类似物咪唑释放融合蛋白即可。该方法仅需一步就可获得很纯的融合蛋白。天然蛋白鲜有寡聚组氨酸区，因而结合到柱子上的基本上只有融合蛋白。

如果这个寡聚组氨酸标签干扰蛋白质的活性怎么办？载体的设计者很周到地提供了去除标签的方法。在多克隆位点前有一段肠激酶（一种蛋白酶，不是真正的激酶）所识别的氨基酸编码区。肠激酶可以将融合蛋白切成寡聚组氨酸标签和目的蛋白两部分。由于肠激酶所识别的位点非常稀有，存在于任意给定蛋白质中的概率可忽略。因此，切开寡聚组氨酸标签时目的蛋白部分不会被酶切。肠激酶切开的蛋白质可再次通过镍柱将寡聚组氨酸片段与目的蛋白质分离开。

λ 噬菌体也可作为表达载体的基础，为此而专门设计的载体有 λgt11。该载体（图 4.21）带有 *lac* 调控区，后接 *lacZ* 基因。多克隆位点位于 *lacZ* 基因内部，因此插入基因的产物是带有 β-半乳糖苷酶先导区的融合蛋白。

λgt11 是构建和筛选 cDNA 文库的常用载体。在前面所讲的筛选实例中，对目的 DNA 序列是用带标记的寡核苷酸或多核苷酸探针检测的，而 λgt11 可直接筛选表达目的蛋白的克隆。实验所需的主要

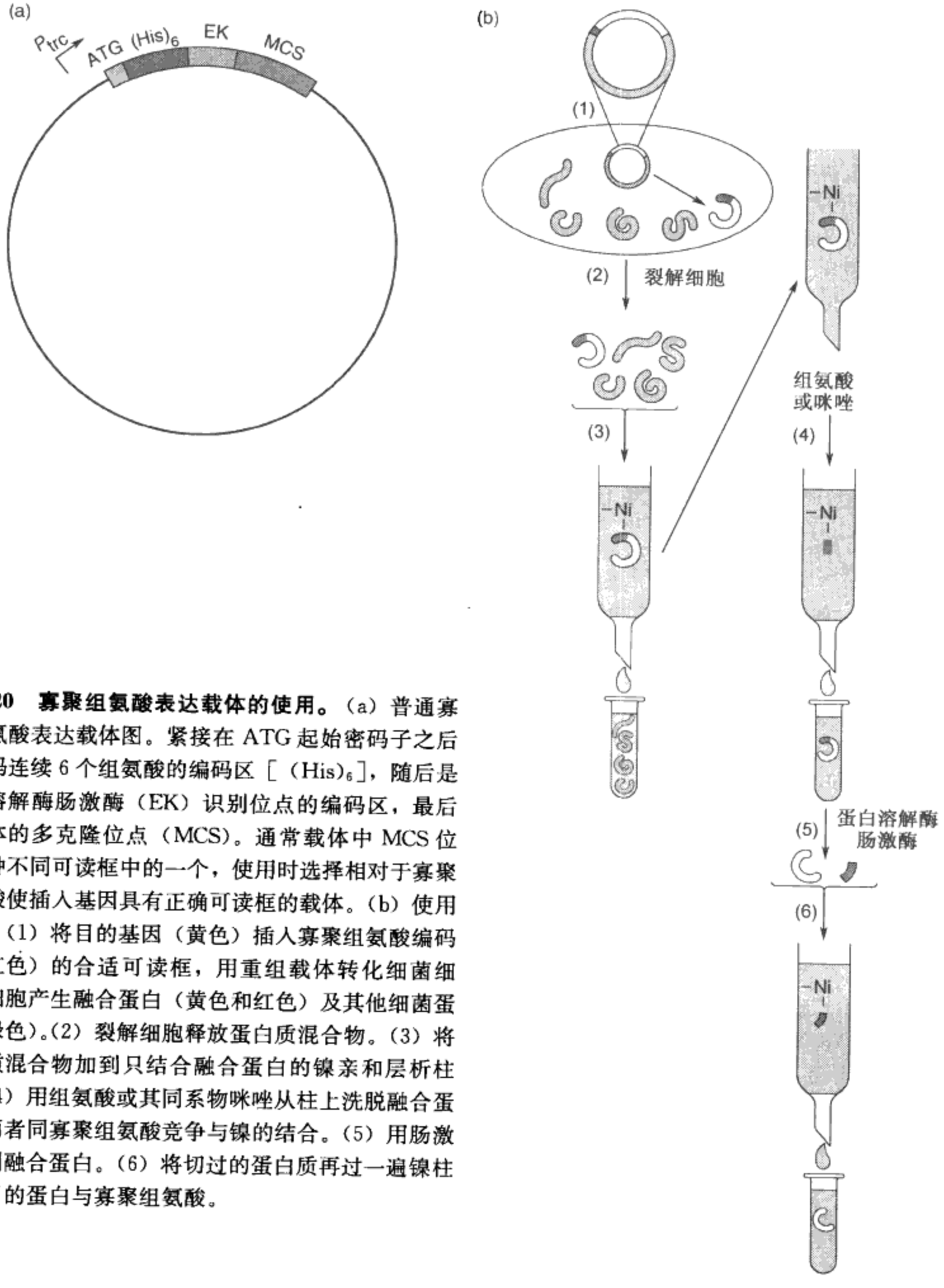


图 4.20　寡聚组氨酸表达载体的使用。（a）普通寡聚组氨酸表达载体图。紧接在 ATG 起始密码子之后是编码连续 6 个组氨酸的编码区［$(His)_6$］，随后是蛋白溶解酶肠激酶（EK）识别位点的编码区，最后是载体的多克隆位点（MCS）。通常载体中 MCS 位于 3 种不同可读框中的一个，使用时选择相对于寡聚组氨酸使插入基因具有正确可读框的载体。（b）使用载体。（1）将目的基因（黄色）插入寡聚组氨酸编码区（红色）的合适可读框，用重组载体转化细菌细胞，细胞产生融合蛋白（黄色和红色）及其他细菌蛋白（绿色）。（2）裂解细胞释放蛋白质混合物。（3）将蛋白质混合物加到只结合融合蛋白的镍亲和层析柱上。（4）用组氨酸或其同系物咪唑从柱上洗脱融合蛋白，两者同寡聚组氨酸竞争与镍的结合。（5）用肠激酶切割融合蛋白。（6）将切过的蛋白质再过一遍镍柱分离目的蛋白与寡聚组氨酸。

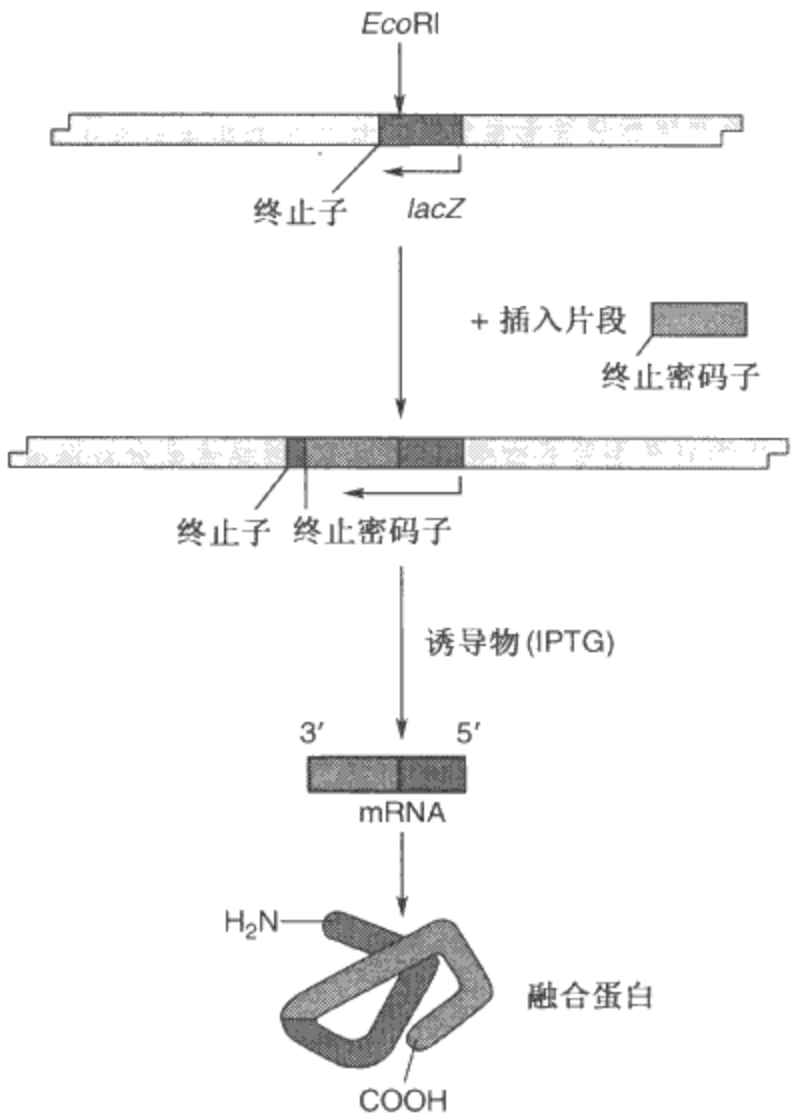


图 4.21　在 λgt11 中合成融合蛋白。要表达的基因插入靠近 *lacZ* 编码区末端的 *Eco*RI 位点，而 *lacZ* 编码区紧连在转录终止子的上游。因此，*lac* 基因在 IPTG 诱导下产生融合 mRNA，其中包含大部分的 β-半乳糖苷酶编码区和紧接其下游的插入编码区。这个 mRNA 在宿主细胞中翻译合成融合蛋白。

成分是用 λgt11 构建的 cDNA 文库和目的蛋白的抗血清。

图 4.22 显示其工作原理。将带有各种 cDNA 插入子的 λ 噬菌体涂平板培养，每个克隆所释放的蛋白质被印迹到支持介质（如纤维素膜）上，然后用抗血清检测。与特定噬菌斑的蛋白质结合的抗体可用标记的葡萄球菌（*Staphylococcus aureus*）蛋白 A 检测。该蛋白质与抗体紧密结合，并标记纤维素膜上的相应斑点。通过放射自显影或磷光图像检测（第 5 章）这个标记，最后将相应的噬菌斑从原始平板中挑出。注意检测的是融合蛋白而不是目的蛋白本身。而且，能否克隆出全长 cDNA 关系并不大。抗血清是抗体混合物，可与蛋白质的不同部位反应，所以，即使部分基因也行，只要它同 β-半乳糖苷酶的编码区在相同的方向和可读框融合表达。

即使得到部分 cDNA 也是有价值的，可以通过 RACE 技术获得全长 cDNA。融合蛋白上的 β-半乳糖苷酶标签有助于融合蛋白在细菌细胞中的稳定，还可利用含有 β-半乳糖苷酶抗体的亲和层析柱很容易地进行纯化。

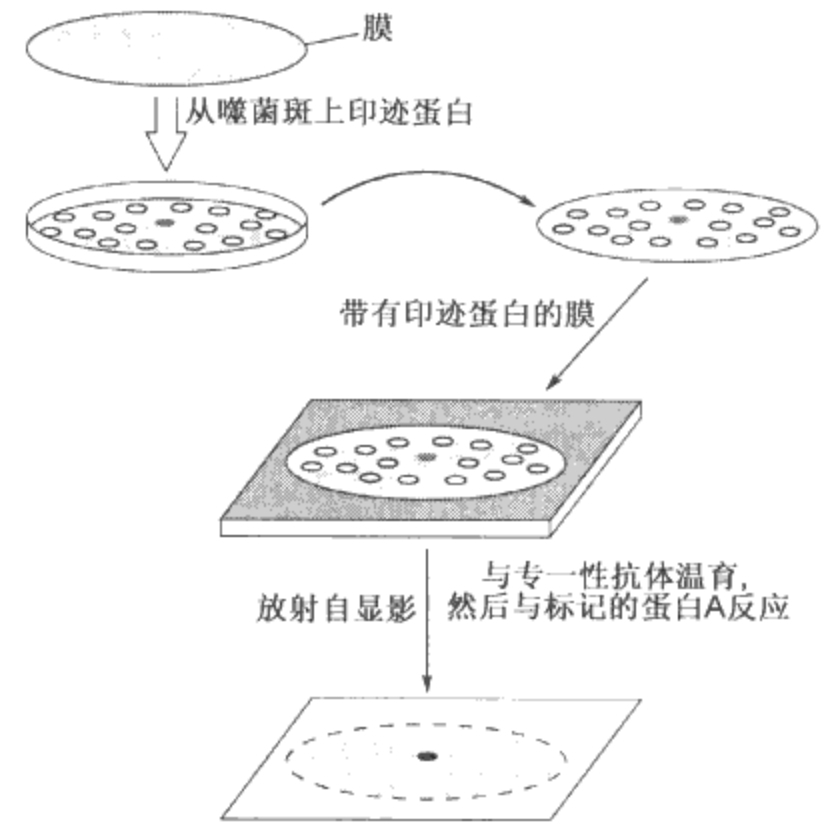


图 4.22　通过抗体筛选检测阳性 λgt11 克隆。用滤膜印迹培养皿中噬菌斑的蛋白质，其中一个克隆（红色）产生了含有融合蛋白的噬菌斑，融合蛋白包含 β-半乳糖苷酶和目的蛋白的一部分。带有蛋白印迹的滤膜与抗目的蛋白的抗体一起孵育，然后用与大多数抗体结合的葡萄球菌蛋白 A 标记。它只能与斑点上阳性克隆相对应的抗原-抗体复合物结合。放置在滤膜上的感光胶片的黑点指示阳性克隆的位置。

小结　表达载体常常生产融合蛋白，其中一部分来自载体的编码序列，另一部分是来自克隆基因本身的序列。许多融合蛋白都具有利用亲和层析简单分离的巨大优点。λgt11 载体产生的融合蛋白可用特异性抗血清在噬菌斑中检测。

真核表达系统　真核基因转化到原核细胞中并没有真正“到家”，即使它们能在原核载体的控制下表达。原因之一是 *E. coli* 细胞有时把克隆的真核基因的蛋白质产物看作外来者而加以清除。另一个原因是原核生物不进行真核生物那样的转录后修饰。例如，在真核细胞中通常要被糖基化的蛋白质在细菌中却表达成一个裸蛋白。这会影响蛋白质的活性或稳定性，或至少影响它对抗体的反应原性。更严重的问题是细菌细胞内环境不像真核细胞那样有利于蛋白质正确折叠。结果常常是错误折叠的无活性基因产物，这就意味着克隆基因在细菌中虽然能以惊人的水平表达，但所形成的产物却是一些高度难溶的无活性的包含体。包含体蛋白是没有用的，除

非能重新折叠并恢复其活性。幸运的是，包含体蛋白的复性经常是可行的。这样的话，包含体反而成了优势，通过简单的离心就能将其与其他几乎所有蛋白质分离开。

为避免克隆基因与宿主之间可能的不相容性，也可在真核细胞中表达基因。在这一过程中，最初的克隆一般在 *E. coli* 中完成，用的是既能在细菌也能在真核细胞中复制的**穿梭载体**（shuttle vector）。然后将重组 DNA 转入所选的真核细胞中。真核细胞酵母很适合这一目的，与细菌一样它也有生长速度快、易于培养的优点，但它又是真核细胞，具有蛋白质折叠和糖基化（加糖）等真核细胞的特征。此外，将克隆基因与编码酵母信号肽的序列连接，可使基因产物分泌到培养基中，这对目的蛋白的纯化有很大的优势。只需通过离心简单地去除酵母细胞，在培养基中留下相对较纯的分泌的基因产物。

酵母表达载体是以正常存在于酵母细胞中的 2μ 质粒（2-micron plasmid）为基础构建的。该质粒提供任何载体在酵母中复制所必需的复制起点。细菌-酵母穿梭载体也含有 pBR322 的复制起点，能在大肠杆菌中复制。此外，酵母表达载体还必须含有一个强的酵母启动子。

另一个十分成功的真核载体来自**杆状病毒**（baculovirus），它是一种感染苜蓿环纹夜蛾幼虫的病毒。这种病毒有一个非常大的环形 DNA 基因组，长约 130kb。主要病毒结构蛋白是多角体蛋白，可在感染细胞中大量合成。据估计，当一只毛虫因感染杆状病毒死亡时，多角体蛋白可达虫体干重的 10%。如此巨大的蛋白质量表明多角体蛋白基因一定非常活跃，而这显然归功于该基因强大的启动子。Max Summers 小组和 Lois Miller 小组利用多角体蛋白启动子分别在 1983 年和 1984 年首次成功地开发出相应载体。从那时起，利用该启动子和其他病毒启动子构建了许多**杆状病毒表达载体**（baculovirus expression vector）。

最佳条件下，杆状病毒表达载体能产生多达 0.5g/L 克隆基因的蛋白质，这确实是非常大的产量。图 4.23 是典型杆状病毒表达系统的工作原理。首先，目的基因被克隆到带有多角体蛋白启动子的载体上（多角体蛋白的编码区已从载体上删除，这不会影响病毒的复制，因为病毒在培养细胞间的传递不需要多角体蛋白）。多数这类载体紧接启动子的下游有唯一的 *Bam*HI 位点，因此可用 *Bam*HI 切割。带有 *Bam*HI 黏端的片段可插入载体，把克隆基因置于多角体蛋白启动子之下。重组质粒（载体＋插入片段）与删除了病毒复制基因和多角体蛋白基因的野生型病毒 DNA 混合，然后用混合物转染培养的昆虫细胞。

因为载体与多角体蛋白基因的侧翼区有广泛的同源性，因而在转染细胞中会发生重组，由此将克隆基因转移到病毒 DNA 中并仍然受多角体蛋白启动子控制。现在这个重组病毒可用来感染细胞了，当细胞进入感染末期时，多角体蛋白启动子处于最活跃状态，就可收获目的蛋白了。而随重组载体一同进入靶细胞的非重组病毒 DNA 由于缺少仅由载体才能提供的必需基因，不会形成有感染性的病毒。

注意，对真核细胞而言，选用转染（transfect）这一术语，而不用对细菌的转化（transform）名词。做这一区别是因为在真核细胞中转化有另一层意思，表示正常细胞转变成癌细胞。为避免与此混淆，使用**转染**表示将新的 DNA 引入真核细胞中。

至少可以用两种方法简便地完成动物细胞的转染。①在磷酸缓冲液中将细胞与 DNA 混合，然后添加钙盐溶液形成磷酸钙沉淀。细胞可吸收这些含有 DNA 的磷酸钙结晶。②将 DNA 与脂质混合，形成**脂质体**（liposome），即内裹 DNA 溶液的小泡。然后脂质体与细胞膜融合，将其 DNA 送入细胞。植物细胞转染通常用**轰击法**（biolistic），即小金属颗粒的表面用 DNA 包被，然后射入细胞内。

小结 外源基因能够在真核细胞中表达。真核表达系统在产生真核蛋白方面比原核表达系统有优势，其中两个最重要的优点是：①真核细胞合成的真核蛋白易于正确地折叠，因而是可溶的而不是集结成不溶性的包含体；②真核细胞产生的真核蛋白能以真核的修饰方式进行修饰（磷酸化、糖基化等）。

其他真核载体

一些著名的真核载体用于其他目的，而非外源基因的表达。例如，酵母人工染色体（yeast artificial chromosome，YAC）、细菌人工染色体（bacterial artificial chromosome，BAC）、P1 噬菌体人工染色体（phage artificial chromosome，PAC）都能接受大片段的外源 DNA，用于像人类基因组计划等大规模测序项目，在这些研究中克隆大片段的 DNA 特别重要。我们将在第 24 章的基因组部分讨论人工染色体。另一个重要的真核载体是 Ti 质粒，它能把外源基因导入植物细胞并确保它们在细胞中的复制。

利用 Ti 质粒将基因导入植物中

利用能在植物细胞中复制的载体可把基因导入植物。普通的细菌载体不能用作这一目的，因为植物细胞不识别细菌启动子和复制起点，而含有 **T-DNA** 的质粒可用作这一目的，该片段来自 **Ti**（肿瘤诱导）质粒。

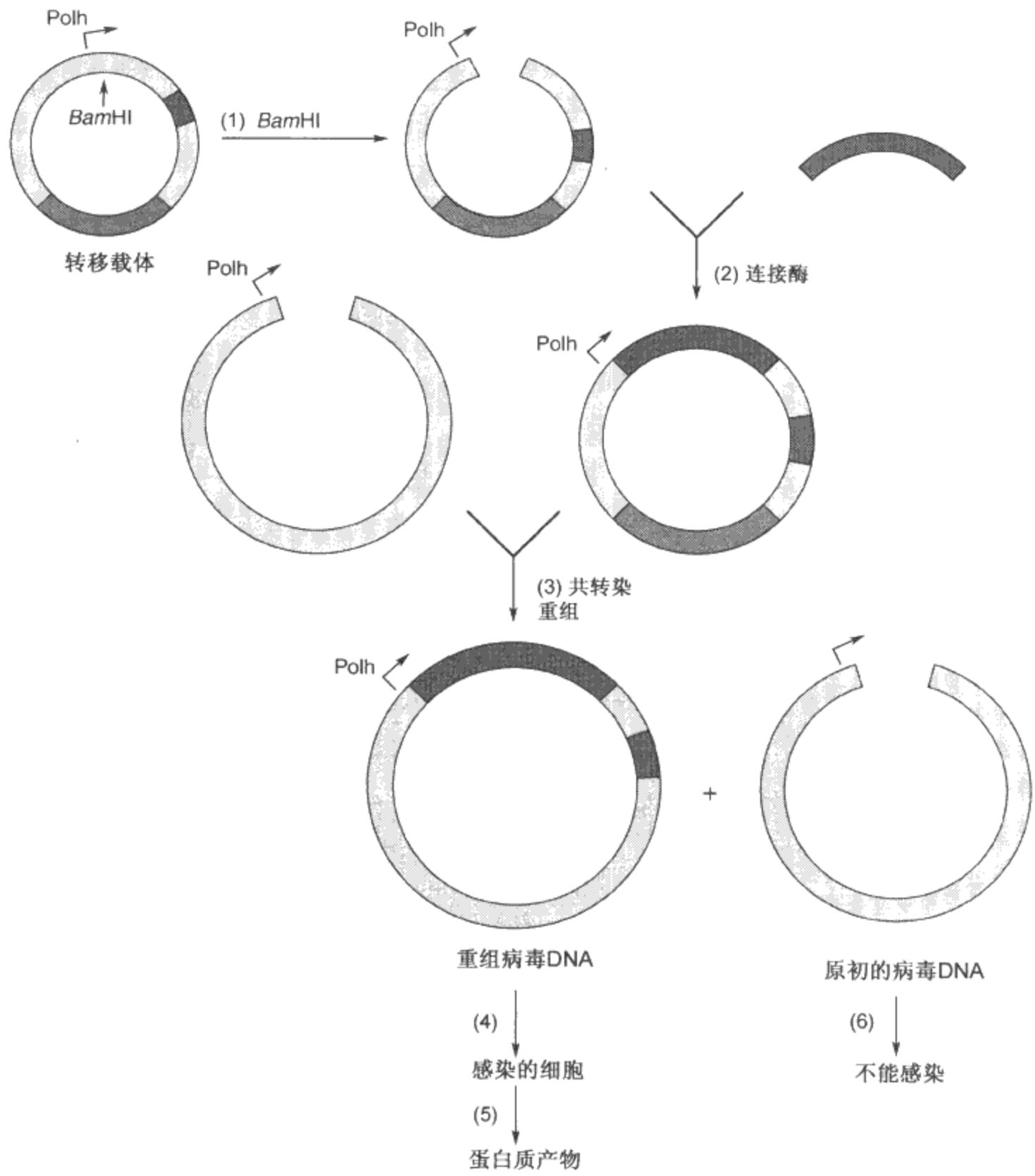
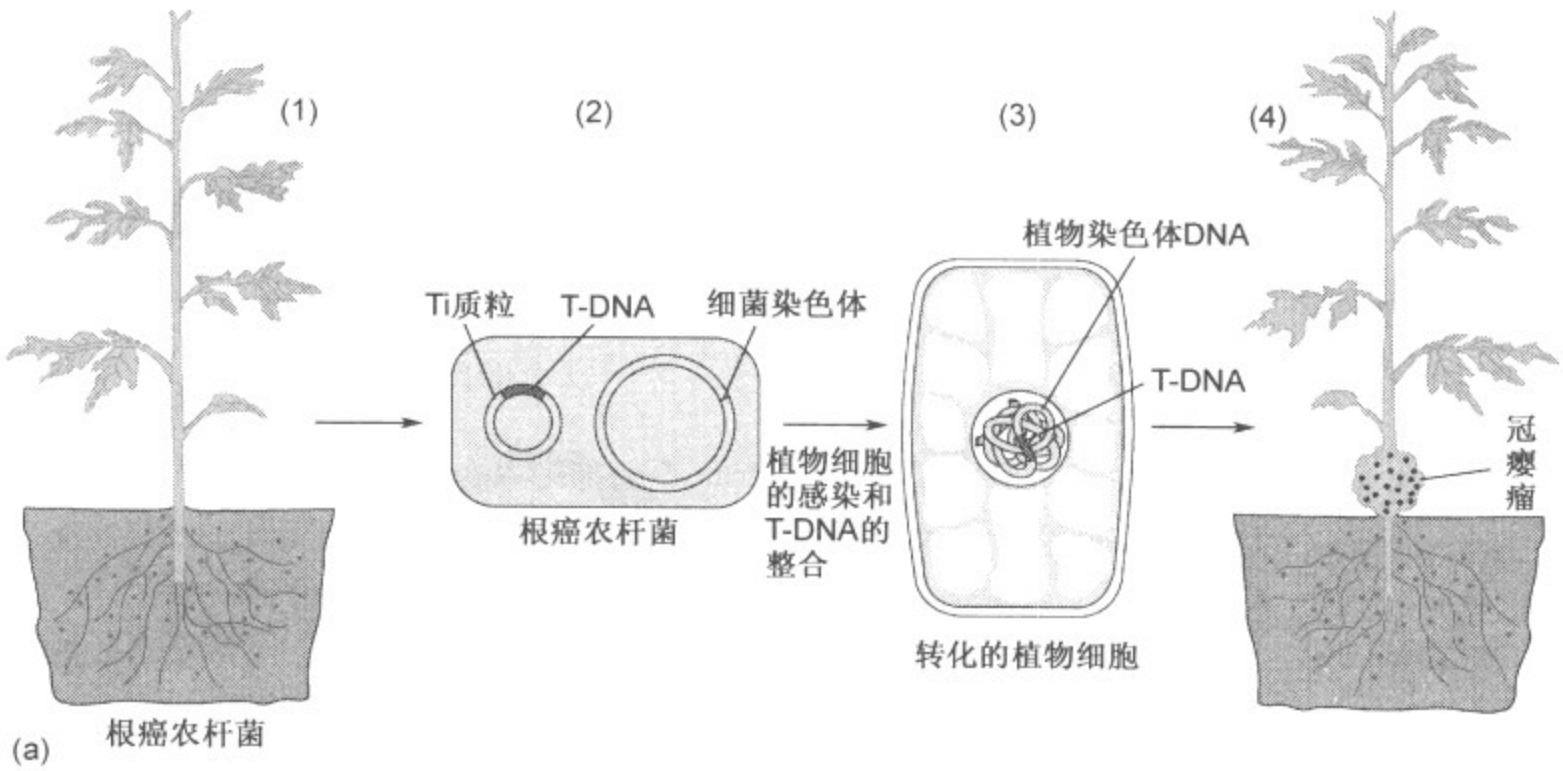


图 4.23　在杆状病毒内表达外源基因。先将要表达的基因（红色）插入杆状病毒转移载体中。转移载体含有强大的多角体蛋白启动子（Polh），两侧的 DNA 序列（黄色）通常包围着多角体蛋白基因，包括病毒复制所必需的基因（绿色）。转移载体中删除了多角体蛋白编码区。细菌载体序列是蓝色的。紧接启动子下游有一个 *Bam* HI 限制酶位点，用于切开载体（步骤 1），通过连接作用（步骤 2）接受外源基因（红色）。在步骤（3）将重组转移载体与已切掉重要基因的线性病毒 DNA 混合，转染昆虫细胞。这个过程称为共转染。两个 DNA 未按比例画，病毒 DNA 实际上几乎是转移载体的 15 倍。在昆虫细胞内，两个 DNA 通过双交换发生重组使目的基因和重要基因一起插入病毒 DNA。结果形成多角体蛋白启动子调控下的目的基因的重组病毒 DNA。最后，在步骤（4）和（5）中用重组病毒感染细胞，收集细胞产生的蛋白质。注意原初的病毒 DNA 是线性的且缺失重要基因，所以不能感染细胞［步骤（6）］。利用这种感染性的缺失，可自动筛选重组病毒，它们是唯一可以感染细胞的部分。

Ti 质粒存在于根癌农杆菌中，在双子叶植物中能引起叫做**冠瘿瘤**（crown gall）的肿瘤（图 4.24）。当这种细菌感染植物时，将自己的 Ti 质粒转移到宿主细胞中，并以 T-DNA 整合到植物 DNA 中，引起植物细胞的异常增殖而产生冠瘿瘤。这对于入侵的细菌很有利，因为 T-DNA 含有编码合成**冠瘿碱**



(a)

(b)

图 4.24　冠瘿瘤。(a) 冠瘿的形成。(1) 根癌农杆菌细胞进入植物的伤口，通常在生长冠或根茎接合处。(2) 根癌农杆菌除含有较大的细菌染色体外还含有 Ti 质粒。Ti 质粒具有一个片段（T-DNA，红色）能促进被感染植物中肿瘤的形成。(3) 根癌农杆菌将其 Ti 质粒转到植物细胞中，Ti 质粒上的 T-DNA 整合到植物基因组上。(4) T-DNA 上的基因指导冠瘿形成，滋养入侵的农杆菌。(b) 切除烟草植株顶部并接种根癌农杆菌形成的冠瘿瘤照片。这种冠瘿瘤是畸胎瘤，可产生正常组织和肿瘤组织。(*Source*：(b) Dr. Robert Turgeon and Dr. B. Gillian Turgeon, Cornell University.)

(opine)酶的基因，这种冠瘿碱对于植物没有价值，但是农杆菌含有分解这种冠瘿碱的酶，将其作为专有能源。

编码合成冠瘿碱的酶（如甘露碱合成酶）的 T-DNA 基因都具有强启动子。利用这个特性可将 T-DNA 连到一个小质粒上，然后把外源基因置于其中一个启动子之下。图 4.25 简述了将外源基因转入烟草获得**转基因植株**（transgenic plant）的过程。从烟草叶片上打下一个小圆片（直径约 7mm），然后将其置于营养培养基中，烟草组织就会沿圆片的边缘生长。接着，向培养基中添加带有重组 Ti 质粒的农杆菌，这些农杆菌就会感染生长的烟草细胞并将克隆的基因导入其中。

当烟草组织沿边缘长出根后就转移到诱导生芽的培养基中。这些小苗会长成完整的烟草植株，其细胞中含有外源基因。该基因能赋予植株新的特性，如抗虫、抗旱、抗病等。

迄今为止，在植物基因工程中最成功的例子之一是“Flavr Savr”番茄的培育。Calgene 公司的遗传学家把一个使果实在成熟期变软的基因的反义拷贝导入番茄。反义基因的 RNA 产物与正常的 mRNA 互补，因而与 mRNA 杂交而阻断该基因的表达，使番茄在成熟时不会变得太软，可以在植株上自然成熟而无需提早采摘然后人工催熟。

其他植物分子生物学家已跨出了更大的步伐，包括：①赋予植物除草剂抗性；②插入病毒衣壳蛋

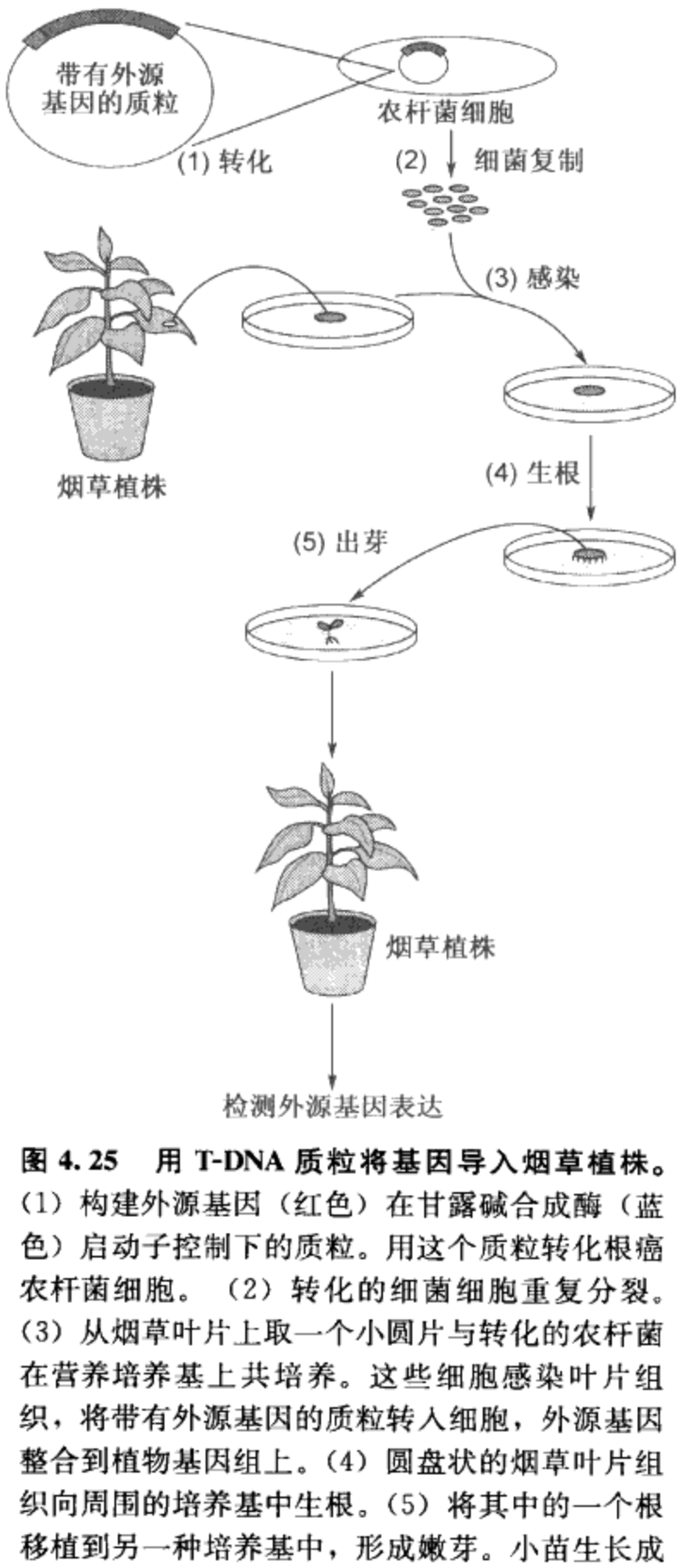


图 4.25　用 T-DNA 质粒将基因导入烟草植株。（1）构建外源基因（红色）在甘露碱合成酶（蓝色）启动子控制下的质粒。用这个质粒转化根癌农杆菌细胞。（2）转化的细菌细胞重复分裂。（3）从烟草叶片上取一个小圆片与转化的农杆菌在营养培养基上共培养。这些细胞感染叶片组织，将带有外源基因的质粒转入细胞，外源基因整合到植物基因组上。（4）圆盘状的烟草叶片组织向周围的培养基中生根。（5）将其中的一个根移植到另一种培养基中，形成嫩芽。小苗生长成转基因烟草，通过外源基因的表达进行检测。

白基因赋予番茄抗病毒性状；③使玉米和棉花产生抗细菌肽；④将萤火虫的萤光素酶基因插入番茄中，这个实验尽管没有实用价值，但的确产生了使植物在黑暗中发光的有趣效果。

小结　生物学家利用植物表达载体（如 Ti 质粒）将所克隆的基因转入植物中，产生了性状改变的转基因生物体。

总结

克隆基因时，必须先将其插入载体，由载体把基因带到宿主细胞中并确保它在细胞中复制。外源基因的插入一般是通过用相同的限制酶切割载体和待插入 DNA 片段，产生相同的黏性末端来完成的。在细菌中用于克隆的载体主要有质粒和噬菌体两种。

在质粒克隆载体中有 pBR322 和 pUC 质粒。pBR322 具有两个抗生素抗性基因和多个独特的用于插入外源基因的限制酶位点。多数限制酶位点都位于一个抗生素抗性基因内，因此可用于直接筛选。用 pUC 质粒和 pBS 噬菌粒进行筛选更容易。这些载体带有氨苄青霉素抗性基因和一个多克隆位点，该位点打断了编码不完整 β-半乳糖苷酶的基因，该基因的产物很容易通过颜色进行筛选。目的克隆具有氨苄抗性，但不能产生有活性的 β-半乳糖苷酶。

有两种很通用的噬菌体克隆载体。首先是 λ 噬菌体，其非必需基因被剔除以便给外源片段的插入腾出空间。其中有些工程噬菌体能容纳长达 20kb 的插入片段。黏粒是噬菌体与质粒载体的交叉载体，可容纳 50kb 的大片段插入。这种特性在构建基因组文库时非常有用。第二类噬菌体载体是 M13 噬菌体，其优点是具有多克隆位点，另一个优点是产生单链重组 DNA，用于 DNA 测序和定点突变。噬菌粒质粒具有单链 DNA 噬菌体的复制起点，可以产生自身的单链拷贝。

设计的表达载体可用于尽可能大规模地生产克隆基因的蛋白质产物。为优化表达水平，细菌表达载体具有强的细菌启动子和细菌核糖体结合位点，该位点在克隆的真核基因中是缺失的。大多数克隆载体（应为表达载体，译者注）是可诱导的，它可避免因外源产物的提早过量合成对宿主菌的毒害作用。表达载体经常会产生融合蛋白，很容易被快速分离出来。真核表达系统的优点是蛋白质产物通常可溶且能以真核细胞的修饰方式被修饰。

用 Ti 质粒等植物载体也可将克隆基因转入植物，这种方法能改变植物的特性。

复习题

1. 参考表 4.1，确定分别由下列限制酶产生的突出端（如果有的话）的长度及其性质（是 5′端还是 3′端）：

a. *Alu*I　　b. *Bgl* II　　c. *Cla*I　　d. *Kpn*I

e. *Mbo*I　　f. *Pvu*I　　g. *Not*I

2. 为什么要把 DNA 连接到载体上进行克隆？

3. 叙述将 DNA 片段克隆到 pBR322 载体的 *Pst*I 位点的过程，如何筛选含有插入片段的克隆？

4. 叙述将 DNA 片段克隆到 pUC 载体的 *Pst*I 和 *Bam*HI 位点的过程，如何筛选含有插入片段的克隆？

5. 叙述将 DNA 片段克隆到 Charon 4 载体 *Eco*RI 酶切位点的过程。

6. 克隆一个 1kb 的 cDNA，本章所讨论的哪些载体适合使用？哪些不适合？为什么？

7. 构建一个 DNA 片段平均长度约 45kb 的基因组文库，本章所讨论的哪种载体最合适？为什么？

8. 构建一个 DNA 片段平均长度大于 100kb 的文库，本章所讨论的哪种载体最合适？为什么？

9. 利用噬菌体载体构建了一个 cDNA 文库。如何从文库中筛选目的基因。叙述使用寡核苷酸探针和抗体探针的方法。

10. 如何从 M13 噬菌体载体上获得克隆的单链 DNA？如果从噬菌粒载体上呢？

11. 图示构建 cDNA 文库的方法。

12. 图示切口平移的过程。

13. 简述用 PCR 扩增一段给定 DNA 片段的方法。

14. 反转录 PCR（RT-PCR）与常规 PCR 的区别是什么？RT-PCR 用于什么目的。

15. 叙述产生一端带有寡聚组氨酸融合蛋白的载体的用途。给出蛋白质纯化流程并说明寡聚组氨酸标签的优点。

16. λ 插入载体和 λ 替换型载体的区别是什么？它们各自的优点分别是什么？

17. 叙述如何利用杆状病毒系统表达克隆基因。

杆状病毒系统相对于原核表达系统的优点是什么？

18. 用何种载体将基因转入植物（如烟草）中？用图示意你的方法。

分析题

1. 以下是待克隆基因的部分氨基酸序列：

Pro—Arg—Tyr—Met—Cys—Trp—Ile—Leu—Met—Ser

a. 选用哪 5 个氨基酸可得到简并性最小的 14 碱基探针用于在文库中探查目的基因？注意，由于第 5 个密码子的简并性，不要用它的最后一个碱基。

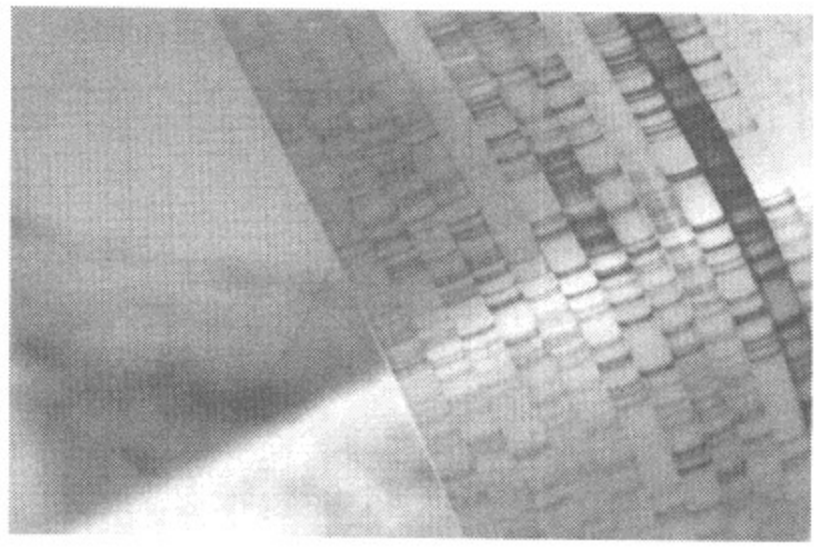
b. 共需要多少个不同的 14 碱基才能够保证探针与克隆基因中相应序列的完全匹配？

c. 如果从所选探针（a 中）向左移一个氨基酸处开始探针，那么需要合成多少种 14 碱基？用遗传密码确定简并数。

2. 把一个新的病毒基因组 DNA 克隆到 pUC18 载体中。将转化子在含有 X-gal 的氨苄青霉素培养皿上涂板，当检测每个质粒（蓝色和白色）中插入片段的大小时，我们惊奇地发现蓝色菌落的质粒含有约 60bp 非常小的插入片段，而来自白色克隆的质粒却未含任何插入片段，请解释这些结果。

翻译　马　纪　　校译　张富春　郑用琏

第5章　研究基因及基因活性的分子工具



DNA 测序胶的自显影图。© *Peter Dazeley/Stone/Getty.*

这一章，我们将了解分子生物学家研究基因结构和功能最常用的技术。这些技术多以克隆基因开始，很多会用到凝胶电泳的方法，也有些要用到示踪标记物，以及核酸杂交法。上一章，我们已经学习了基因克隆技术，现在简要了解一下分子生物学研究的三大支撑技术：凝胶电泳的分子分离、标记性示踪物和杂交。

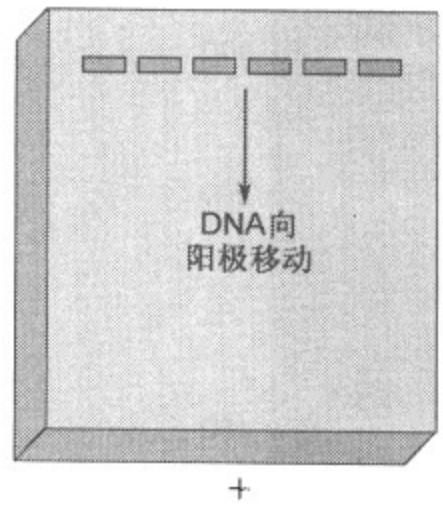
5.1　分子的分离

凝胶电泳

在分子生物学研究中分离蛋白质与核酸是最常见的事。例如，从细胞粗提物中纯化一种特殊的酶，以便利用或研究其性质，或者纯化一种经酶促反应产生或修饰的 RNA 或 DNA 分子，或者只是想分离一系列 RNA 或 DNA 片段。本节将讨论几种最常用的分子分离技术，包括蛋白质和核酸的凝胶电泳、离子交换层析、凝胶过滤层析。

凝胶电泳（gel electrophoresis）　凝胶电泳可用于核酸和蛋白质的分离。首先从 DNA 的凝胶电泳开始。在这个技术中，要先制备一个带槽的凝胶（图 5.1）。方法是将热的（液体的）琼脂糖溶液倒进一个插有“梳子”的浅盒子里。待胶凝固后拔去梳子留下长方形的小孔或槽。向胶孔中加入少许 DNA，在中性 pH 下，接通电源使电流通过凝胶。DNA 因骨架上的磷酸而带负电荷，因此向凝胶另一端的阳极（正极）迁移。凝胶能分离不同大小 DNA 片段的奥秘在于摩擦力，小分子 DNA 由于与溶剂和凝胶分子间的摩擦力小，因此迁移较快；而大分

(a)



(b)

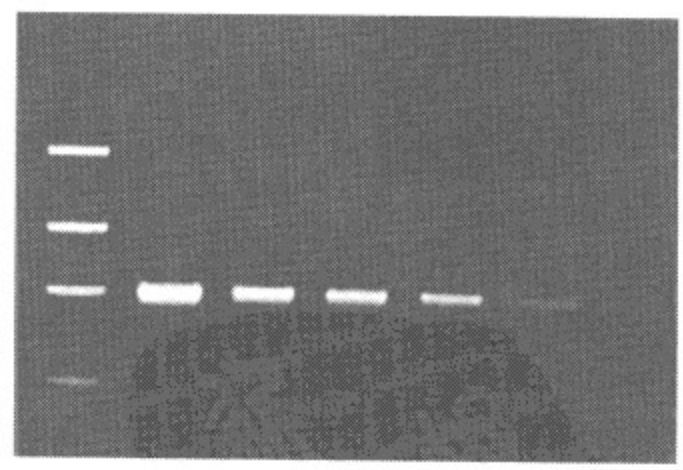


图 5.1　DNA 凝胶电泳。（a）该方法的图解，这是琼脂糖（从海藻中分离的物质，主要成分是琼脂）制备的水平凝胶。琼脂在高温下熔化，冷却后形成凝胶。凝胶熔化时插入一把“梳子”，冷却后拔出梳子形成狭缝或槽（橙色部分）。然后将 DNA 加入槽中，在凝胶中通入电流。由于 DNA 是酸性的，在中性 pH 下带负电荷，所以会朝阳极（正极）电泳或移动。（b）凝胶电泳后的照片，DNA 片段显示为明亮的条带。DNA 与在紫外光下发橙色荧光的染料结合，但是本张照片中的条带显示为粉红色。（*Source*：（b）Reproduced with permission from Life Technologies, Inc.）

子 DNA 受到的摩擦力相对较大，因此迁移得较慢。电流根据 DNA 片段大小将它们分离开，最大的靠近顶凝胶顶端，而最小的靠近凝胶底部。最后用荧光染料对 DNA 染色并将凝胶置于紫外灯下检测。图 5.2 显示了用这种方法对已知大小的噬菌体 DNA 片段分析的结果。将片段的迁移距离与其分子质量（或碱基对的数目）的对数值作图。任何未知大小的 DNA 都能与标准 DNA 片段同时电泳，只要其落在标准 DNA 片段的范围之内，就能估测其大小。例如，图 5.2 中一个迁移率为 20mm 的 DNA，其片段大小约 910bp。同样，也可用此方法对不同大小的 RNA 片段进行电泳。

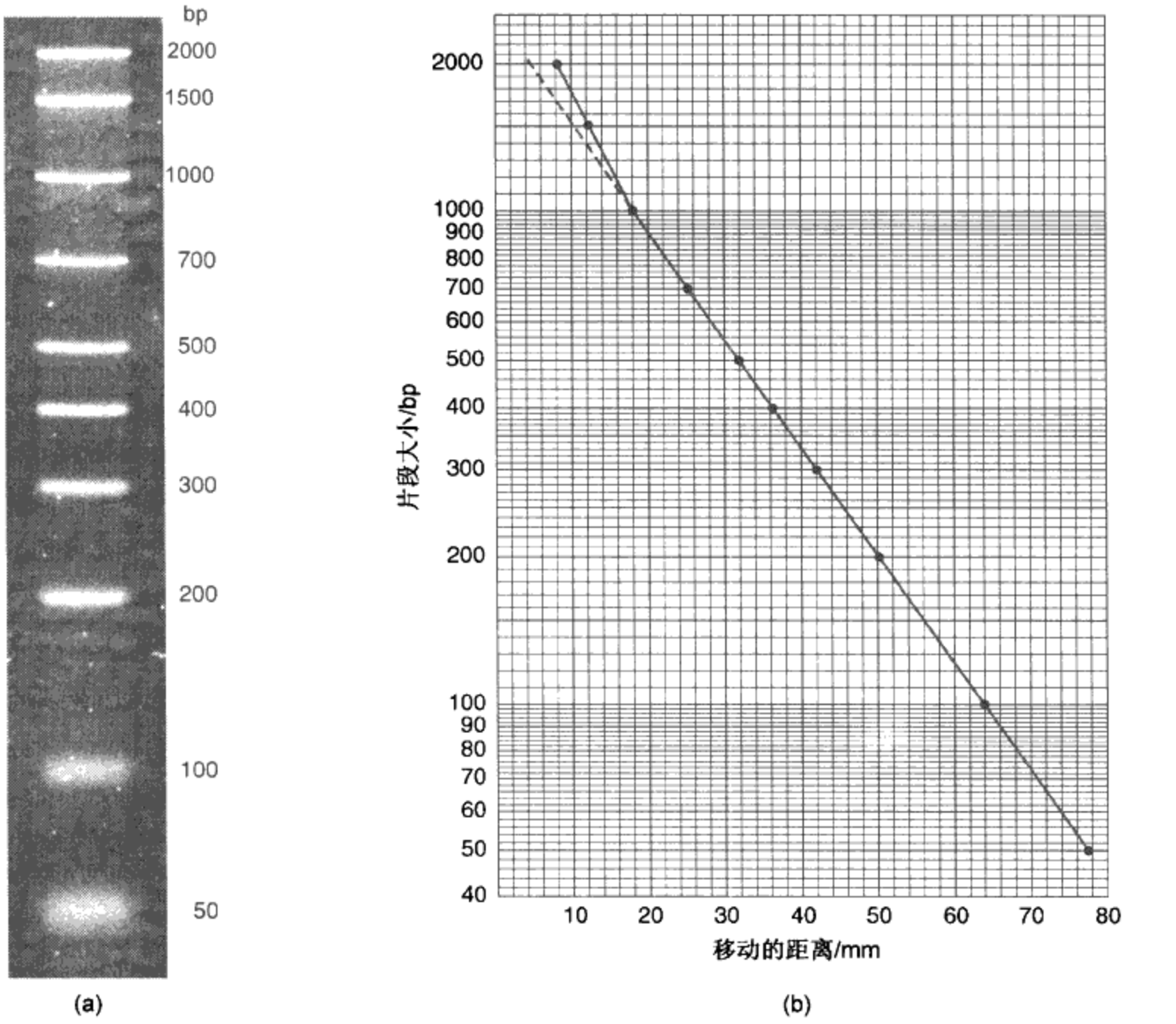


图 5.2　DNA 片段大小的凝胶电泳分析。（a）商业上制备的 DNA 片段电泳后的着色凝胶照片，在彩色照片中应该显橙色的条带在黑白照片中经橙色滤镜过滤后显白色。条带大小（bp）在右侧给出。图已被放大，故条带的迁移距离比实际看上去要大些。（b）DNA 片段的迁移距离与其碱基对大小的曲线图。纵轴是对数的而不是线性的，DNA 片段电泳的速度（迁移率）与其大小的对数成反比。但要注意大片段 DNA 会偏离这种比例，显示为实线（实际结果）与虚线（理论结果）的差异，说明常规电泳在检测非常大的 DNA 时具有局限性。（*Source*：(a) Courtesy Bio-Rad Laboratories.）

问题解答

问题 1

下面是 0.3～1.2kb 大小的双链 DNA 片段的电泳结果图。请根据下页图示回答下列问题：

a. 迁移到 16mm 处的 DNA 片段的大小是多少？

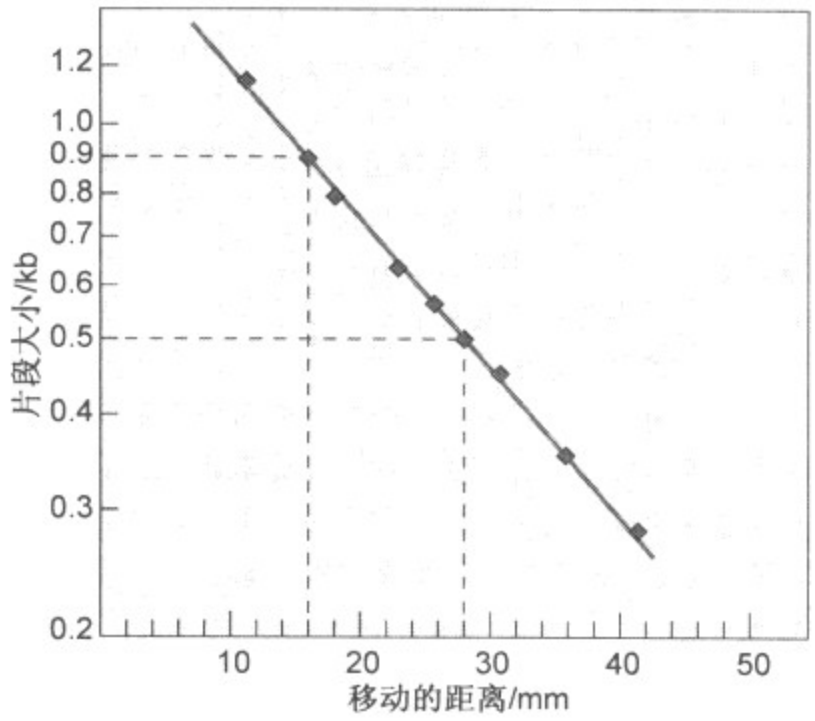
b. 0.5kb DNA 片段的迁移距离约为多少？

解答

a. 从 x 轴的 16mm 处画一条垂直虚线与实验线相交，再从交点处画一条水平虚线与 y 轴相交，如图虚线与 y 轴的交点是 0.9kb，说明在实验中迁移 16mm 的 DNA 片段的大小是 0.9kb（900bp）。

b. 从 y 轴的 0.5kb 点处画一条水平虚线与实验线相交，再从交点画一条垂直虚线与 x 轴相交，如图虚线与 x 轴的交点是 25mm，说明在实验中 0.5kb DNA 片段的迁移距离是 25mm。

用凝胶电泳检测大分子 DNA 需要特殊的技术。原因之一是如果 DNA 太大，其大小的对数值与迁



移的距离会严重偏离线性关系，图 5.2b 的左上部出现了偏离的迹象。另一个原因是双链 DNA 是细长的刚性棒状结构，越长就越脆弱。大分子 DNA 即使在温和条件下也很容易断裂，如在烧杯中搅拌或移液操作等产生的剪切力都足以将其打断。形象地说，可以把 DNA 想象成生的意大利面条，假如只有 1～2cm，动作粗放点不会损伤它，但是如果较长，断裂就是不可避免的了。

尽管有很多困难，分子生物学家还是发明了一种新的凝胶电泳，可分离长达几百万个碱基对（megabase，Mb）的 DNA 分子，且能保证其大小的对数值与迁移率的相对线性关系。这种方法不是采用恒定电流而是用脉冲电流：正向的是相对较长的电脉冲，反方向或旁侧方向的是短脉冲。这种**脉冲场凝胶电泳**（pulsed-field gel electrophoresis，**PFGE**）对于测定像酵母染色体这样大的 DNA 时非常有用。图 5.3 显示的是脉冲场凝胶电泳分离酵母染色体的结果，16 条可见条带代表含有 0.2～2.2Mb 碱基对的染色体。

电泳也常用于蛋白质的分离，只是其凝胶介质通常由聚丙烯酰胺制成，因此称为**聚丙烯酰胺凝胶电泳**（polyacrylamide gel electrophoresis，**PAGE**）。为了确定某种复杂蛋白质的多肽组成，必须处理蛋白质，以便多肽链或亚基独立地电泳移动。通常利用去垢剂**十二烷基磺酸钠**（sodium dodecyl sulfate，SDS）处理蛋白质，使各个亚基**变性**（denature）不再相互结合。SDS 有两个附加优点：①使所有多肽的表面都覆盖上负电荷，从而在电泳时都会往阳极泳动；②遮蔽蛋白亚基的原有电荷，使各亚基在电泳时按分子质量大小而不是原有电荷迁移。小分子多肽容易穿过凝胶上的小孔，因而迁移得较快，而大分子多肽则相对较慢。研究者也经常用还原剂打开亚基间的共价键。

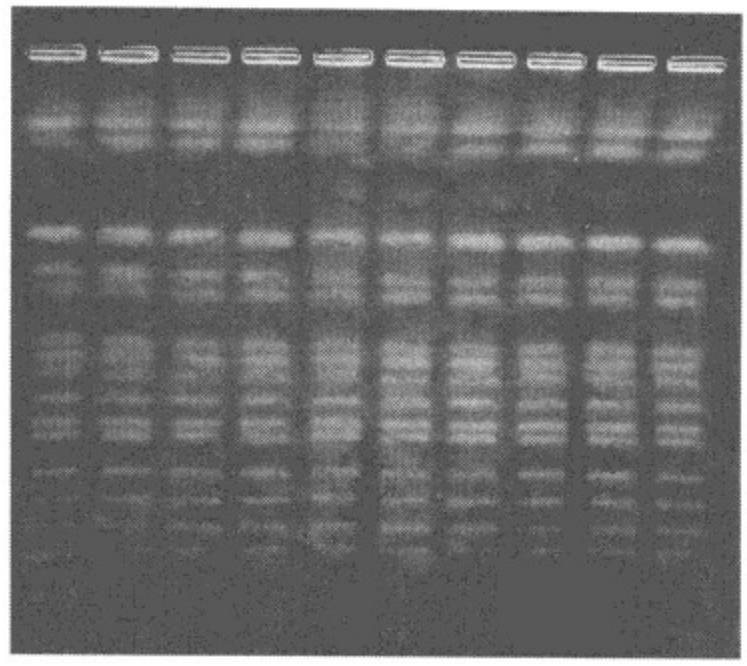


图 5.3　酵母染色体的脉冲场电泳。相同的酵母染色体在 10 个平行泳道中电泳，用溴化乙锭染色。各个条带显示染色体大小从 0.2Mb（底部）到 2.2Mb（顶部）。凝胶的原大小为 13cm 宽 12.5cm 长。（*Source*：Courtesy Bio-Rad Laboratories/CHEF-DR（R）II pulsed-field electrophoresis systems.）

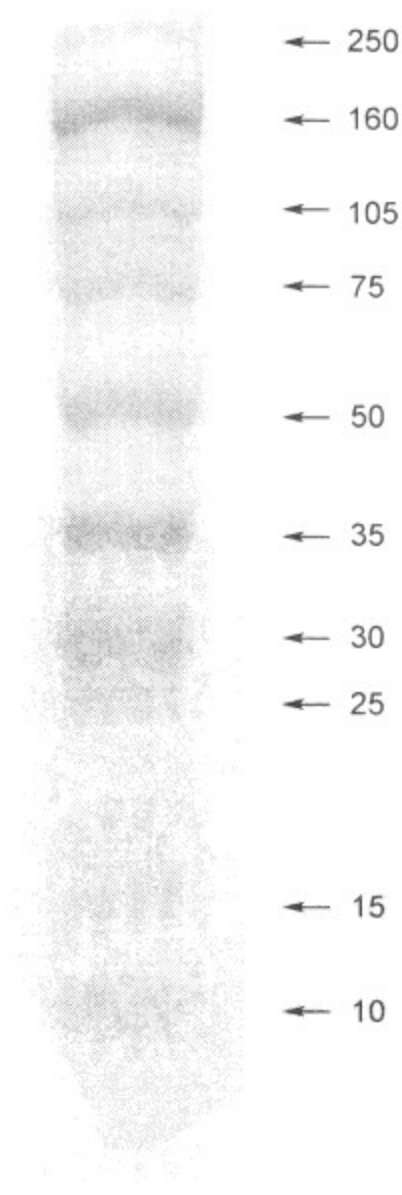


图 5.4　SDS-PAGE。对多肽链染色后进行 SDS-PAGE，右侧显示的是肽链的分子质量。通过染色可以在电泳中和电泳结束后观察每条多肽。（*Source*：Courtesy of Amersham Pharmacia Biotech.）

图 5.4 显示的是 SDS-PAGE 对一组多肽链的分离结果，每个多肽链都结合了一种染料，以便在电泳时能看到。通常的做法是在电泳后用染料如考马斯亮蓝对所有多肽染色。

小结 通过凝胶电泳可把各种分子质量的 DNA、RNA 和蛋白质分离开来。在核酸电泳中常用的凝胶是琼脂糖，而蛋白质电泳中常用的是聚丙烯酰胺。SDS-PAGE 通常是根据分子质量的大小分离多肽。

双向凝胶电泳

SDS-PAGE 能很好地分离多肽，但有时多肽混合物太复杂了，需要用更好的方法把它们分离开。例如，我们想把某一特定时间某种特定细胞中成千上万条多肽分离开，这在目前的蛋白质组学（分子生物学的一个分支）中经常要做，有关蛋白质组学我们将在第 24 章中讨论。

为了提高一维 SDS-PAGE 的分辨率，分子生物学家建立了双向方法。在第 19 章将介绍一种简单的方法，首先在一个 pH 和凝胶浓度下进行单向非变性（无 SDS）凝胶电泳，然后在另一个 pH 和凝胶浓度下进行第二向电泳。由于蛋白质所带的净电荷随溶液 pH 的变化而变化，因此在不同的 pH 下电泳的迁移率不同。在不同聚丙烯酰胺浓度下根据其分子质量大小，蛋白质也会以不同的迁移速率电泳。但是由于缺乏去垢剂，不能将组成复杂的蛋白质的各个肽链分开，因此用这种方法不能分析单个多肽。

另一种常用的更有效的方法是**双向凝胶电泳**（two-dimensional gel electrophoresis），其实它所涉及的范围比其名字所包含的内容更多一些。第一步，蛋白质混合物在一个细的管状凝胶中电泳，凝胶中含有一种两性电解质（ampholyte），可在毛细管的两端之间形成 pH 梯度。带负电荷的蛋白质分子在管状凝胶中朝阳极迁移，到达其**等电点**（isoelectric point），即蛋白质的净电荷为 0 的 pH。由于没有净电荷，蛋白质不再向阳极或阴极迁移，这一步叫做**等电聚焦**（isoelectric focusing），使蛋白质在凝胶中相应的等电点聚集。

第二步，将管中的凝胶取出，置于平板凝胶的顶端进行常规的 SDS-PAGE。这时，经等电聚焦初步分离的蛋白质将根据其大小通过 SDS-PAGE 进一步分离。图 5.5 显示在有或无安息香酸条件下生长的 *E. coli* 总蛋白的双向电泳结果。无安息香酸条件下 *E. coli* 蛋白用红色荧光染料 Cy3 染色，有安息香酸条件下 *E. coli* 蛋白用蓝色荧光染料 Cy5 染色。从两组蛋白质各自的和混合的双向凝胶电泳图可以看出两种条件下哪些蛋白质含量更丰富，哪些蛋白质在两种条件下均占优势。

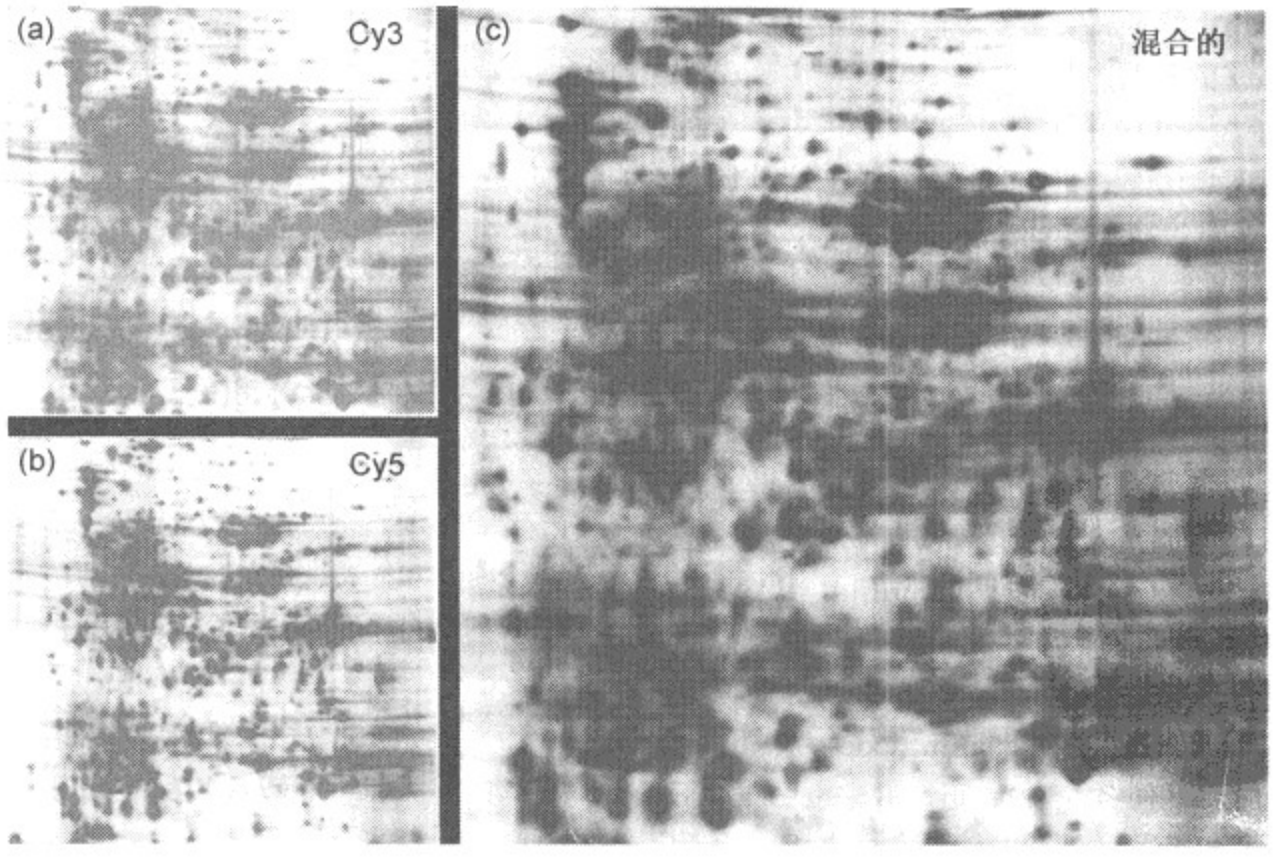


图 5.5 双向凝胶电泳。在有或无安息香酸条件下培养 *E. coli*。然后用红色荧光染料 Cy3 对无安息香酸条件下培养的 *E. coli* 细胞裂解液染色，蛋白质被染成红色。用蓝色荧光染料 Cy5 对有安息香酸条件下培养的 *E. coli* 细胞裂解液染色，蛋白质被染成蓝色。最后进行双向凝胶电泳，（a）无安息香酸条件下细胞中的蛋白质，（b）有安息香酸条件下细胞中的蛋白质，（c）两组蛋白质的混合物。在（c）中只在无安息香酸条件下积累的蛋白质显红色荧光，在有安息香酸条件下积累的蛋白质显蓝色荧光，在两种情况下都积累的蛋白质既显红色又显蓝色，所以最后呈蓝紫色或黑色。（*Source*：Courtesy of Amersham Pharmacia Biotech.）

小结 双向凝胶电泳通过第一向等电聚焦和第二向 SDS-PAGE 能很好地达到分离多肽的效果。

离子交换层析

色谱（层析）（chromatography）最早是指在纸上［**纸层析**（paper chromatography）］将有色物质分离后形成的图案。目前有多种不同的层析技术用来分离生物物质。**离子交换层析**（ion-exchange chromatography）是利用树脂分离不同电荷的物质。例如，DEAE-Sephadex 层析用的是带正电荷的二乙氨乙基（diethylaminoethyl，DEAE）离子交换树脂，其正电荷基团吸引包括蛋白质在内的带负电荷的物质，所带负电荷越多，结合就越牢固。

我们将在第 10 章以 DEAE-Sephadex 层析为例介绍如何利用离子交换层析分离 3 种形式的 RNA 聚合酶。首先，制备 DEAE-Sephadex 浆液并装柱，当树脂沉积后加入含 RNA 聚合酶的细胞粗提物。最后，通过逐渐升高洗脱液的盐离子强度（或盐溶液浓度），将结合在树脂上的物质**洗脱**（elute）下来或去除。设置盐浓度梯度的目的是用盐溶液中的阴离子与蛋白质竞争树脂上的离子结合位点，从而将蛋白质逐一洗脱下来。这就是为什么将其称为离子交换层析的原因。

随着洗脱缓冲液离子强度的升高，用组分收集器（fraction collector）收集洗脱液。该装置工作时每次将一个试管置于层析柱下收集一定体积的溶液，然后移到旁边，由下一个试管进入并收集组分。最后，分析每个洗脱组分中目的物的含量。如果目的物是一种酶，需要分析各洗脱组分中该酶的特殊活性。另外，检测每组洗脱液的离子强度有助于确定洗脱目的蛋白所需要的盐浓度。

也可以用带负电荷的交换树脂分离带正电荷的物质，包括蛋白质。例如，磷酸纤维素常用于阳离子交换层析分离蛋白质。注意，蛋白质不一定必须带有净正电荷才能与磷酸纤维素等阳离子交换树脂结合，大多数蛋白质带净负电荷。但是，某蛋白质如果有一个显著的正电荷中心，则仍能结合到阳离子交换树脂上。图 5.6 是假定用离子交换层析法分离酶的两种构象的实验结果。

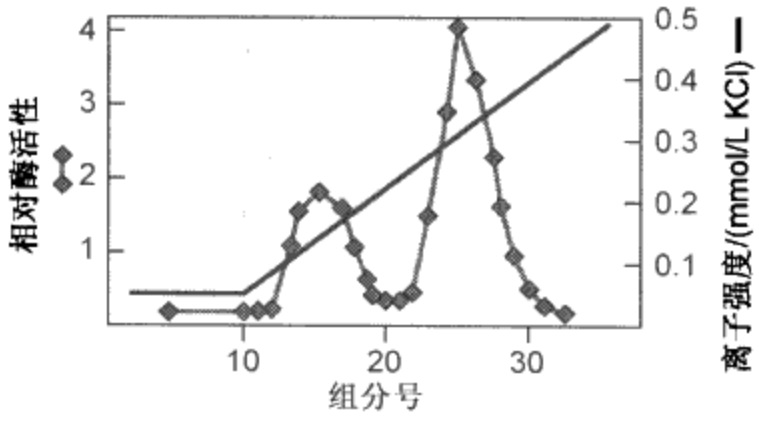


图 5.6 离子交换层析。先将细胞提取液加样到离子交换柱上，提取液含有两种不同构象的酶，然后不断增加洗脱缓冲液的离子强度，同时收集组分（本例共收集了 32 个）。分析每个组分的酶活性（曲线）和离子浓度（直线）并作图。酶的两种构象被清楚地分离开了。

小结 离子交换层析可根据所带电荷的多少分离蛋白质等物质。DEAE-Sephadex 等带正电荷的树脂常用在阴离子交换层析中，而磷酸纤维素等带负电荷的树脂常用在阳离子交换层析中。

凝胶过滤层析

标准的蛋白质生化分离技术通常需要一个以上的步骤，但是宝贵的蛋白质在每一步都会有损失，因此尽量减少蛋白质分离的步骤十分重要。实现这一目的的一种方法是制定一个使每步分离都能利用目的蛋白的不同性质的实验策略。如果第一步用的是阴离子交换层析，第二步是阳离子交换层析，那么第三步就应该根据电荷以外的其他特性来分离蛋白质，蛋白质大小显然是下一个选择。

凝胶过滤层析（gel filtration chromatography）是一种基于分子自身大小（physical dimension）进行分离的方法。凝胶过滤介质，如 Sephadex 是不同大小的多孔珠子，好比“威浮球”的带孔空心塑料球。将柱子中充满带孔的微型塑料球，当不同大小的分子通过柱子时，小分子很容易进入威浮球的小孔内（磁珠上的小孔），因此流过柱子的速度较慢。而大分子不能进入珠子内，因而较快速地流过柱子，出现在所谓的**外水体积/空柱体积**（void volume）即珠子周围的缓冲液体积，不包括珠子内的液体。中等大小的分子选择性进入一些孔径大小适合的珠子，因此迁移率居中。这样，大分子首先从柱中流出，小分子最后流出。具有不同大小的各类树脂可用来分离不同大小的分子。图 5.7 解释了这一方法。

小结 凝胶过滤层析柱以多孔树脂填充，使小分子物质进入树脂孔而大分子不能进入。这样，小分子物质通过柱子的速率慢，而大分子物质则会相对较快地流出柱子。

亲和层析

最有效的分离技术之一是**亲和层析**（affinity chromatography），树脂是含有与目的分子有很强专一性的亲和试剂。例如，树脂可耦合能够识别特殊蛋白的抗体，或含有酶底物的无活性同系物。在后一种情况下，酶与同系物紧密结合但不发生反应。当杂蛋白因为与亲和试剂无（或很弱）亲和性而流

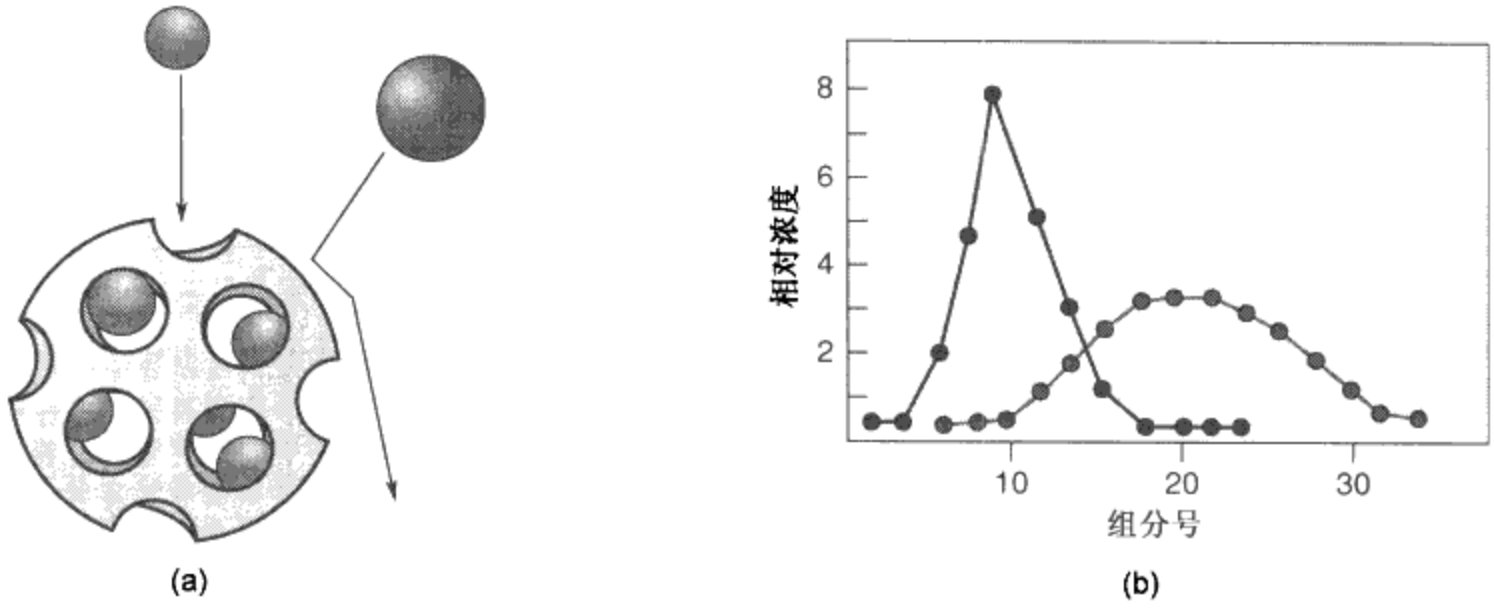


图 5.7　凝胶过滤层析。(a) 原理。以树脂珠代表多孔“威浮球”（黄色）。大分子（蓝色）不能进入珠子，因此它们局限在树脂珠外相对小的洗脱体积中，从柱中快速洗出。而小分子（红色）能进入树脂孔，需要较大的洗脱液体积以及较长的时间才能从柱中洗出。(b) 实验结果。将 (a) 中的大小分子混合物加到柱子上，用缓冲液洗脱。收集组分并分析每个组分中大分子（左峰）与小分子（右峰）的浓度。和预期一样，大分子早于小分子洗脱出来。

过柱子后，利用与目的分子竞争结合亲和试剂的溶液从柱子上洗脱出目的分子。例如，用酶的同系物溶液，溶液中的同系物与树脂上的同系物竞争对酶的结合，使酶从柱上洗脱下来。

亲和层析的效果取决于树脂上的亲和试剂与要纯化分子间结合的专一性。设计一个一步纯化蛋白质的亲和层析方案是可能的，因为该蛋白质是细胞中唯一能与亲和试剂结合的物质。在第 4 章我们已看到一个很好的例子：用镍柱纯化带有寡聚组氨酸的蛋白质。由于细胞中所有其他的蛋白质都是天然的，因此都不带有寡聚组氨酸标签，所以带标签的蛋白质就是唯一能结合亲和试剂镍的蛋白质。然后用镍溶液洗脱蛋白质，得到蛋白质-镍混合物，再用组氨酸同系物咪唑（imidazole）结合到镍柱上，干扰亲和试剂与目的蛋白的结合。

当要纯化的目的分子（即带有寡聚组氨酸标签的蛋白质）是唯一与亲和树脂结合的蛋白质时，甚至都不需要层析柱了。只需简单地将树脂与细胞提取物混合，离心沉淀树脂，弃去剩下的溶液［**上清**（supernatant）］留下沉淀在离心管底部的目的蛋白与树脂。用缓冲液洗涤沉淀后，目的蛋白就可以从树脂中释放出来（如果用的是镍树脂，则用咪唑溶液），然后离心，再次沉淀树脂，这一次目的蛋白存在于上清中，可收集保存。这个方法比传统的层析法更简单快捷。

小结　亲和层析是一种有效利用与目的分子有很强专一性的亲和试剂进行分离的纯化技术。目的分子与耦合有亲和试剂的柱子结合，而所有或绝大多数其他分子因不能结合而流出柱子。然后用干扰这种特异性结合的溶液将目的分子从柱上洗脱下来。

5.2　标记性示踪物

至今，“标记性”实际上仍然与“放射性”同义，因为放射性示踪物已经使用了数十年而且容易检测。放射性示踪物可以检测到极微量的物质，这对于分子生物学研究来讲十分重要，因为在典型的分子生物学实验中被测物质的含量通常极微，如检测转录反应中的某种 RNA 产物，我们可能得测定低于 1pg（1g 的兆分之一或 10^{-12} g）的 RNA 量。由于紫外吸收或染色等方法的灵敏度有限，用这些方法直接检测如此微量的 RNA 是不可能的。然而，如果该 RNA 具有放射性，使用高灵敏度的放射性检测装置就能很容易地对微量 RNA 进行检测。现在，我们来学习一下分子生物学家偏爱的检测放射性示踪物的几种方法：放射性自显影、磷屏成像和液体闪烁计数。

放射性自显影

放射性自显影（autoradiography）是使用照相感光乳剂检测放射性复合物的一种方法。分子生物学家偏爱的感光乳剂是 X 线胶片。如图 5.8 所示，凝胶电泳分离放射性标记的 DNA 片段，然后让凝胶与 X 线胶片接触并一起置于暗室中数小时甚至数天。从 DNA 条带发出的放射性就像可见光一样可使 X 线胶片曝光。胶片冲洗之后，与凝胶上的 DNA 条带对应的黑色条带就显现出来。实际上是 DNA

条带给自己拍了照，这就是为什么将这种技术称为放射自显影的原因。

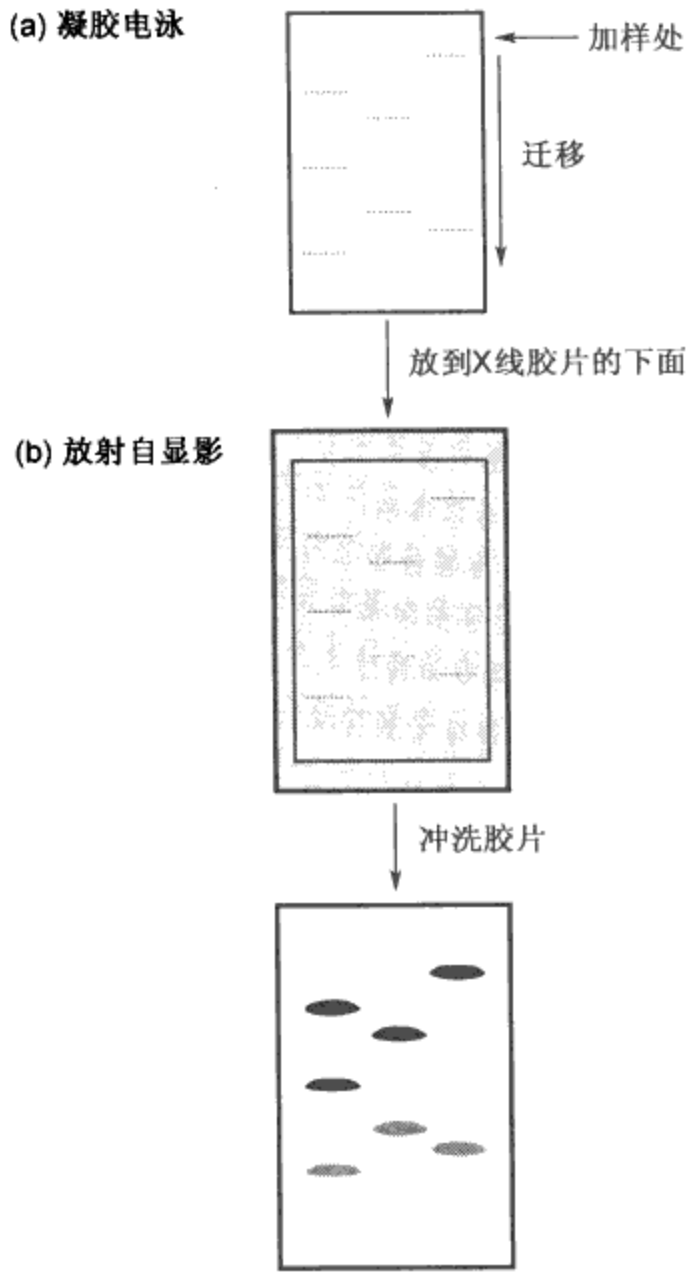


图 5.8　放射自显影。（a）凝胶电泳。在凝胶的 3 个平行泳道电泳放射性 DNA 片段，根据 DNA 片段大小采用琼脂糖或聚丙烯酰胺凝胶。在这一步 DNA 是看不到的，这里用虚线表示它们的位置。（b）放射自显影。将 X 线胶片铺在凝胶上并放置若干个小时，如果放射性比较弱可延长至几天。最后冲洗底片观察哪些地方经射线曝光，由此显示 DNA 条带在凝胶上的位置。本例中，大的且迁移缓慢的条带放射性最强，因此放射自显影中相应的条带最黑。

采用**增感屏**（intensifying screen）可提高放射性自显影的灵敏度，至少是对^{32}P的灵敏度。这是一种涂有荧光复合物的屏板，荧光复合物在低温时能被 β 电子激发（β 电子是分子生物学常用的放射性同位素^{3}H、^{14}C、^{35}S和^{32}p等发出的射线）。因此，照相底片的一面放上具有放射性的凝胶（或其他介质），另一面放上增感屏。β 电子就可直接曝光底片，若没有增感屏则大多数 β 电子会直接穿过底片而丢失。当这些高能电子撞击增感屏时可产生荧光，这种荧光能被底片检测到。

怎样对一个 DNA 片段的放射性进行精确定量呢？可以观察放射自显影胶片上条带的亮度进行粗测，也可使用**光密度计**（densitometer）扫描放射自显影图片，进行较精确的测定。这种装置能使光束穿过样品（这里是一张放射自显影照片），然后检测样品的光吸收。如果条带很黑就会吸收大量的光线，则光密度计会记录一个高的光吸收峰（图 5.9）；如果条带比较弱，则大部分光会透过条带，光密度计记录到小的光吸收峰。通过测量每个峰的面积就可估算出每个条带的放射性。当然，这仍是间接检测放射性的方法，要想真正精确读出每个条带的放射性，可以用磷屏成像仪扫描凝胶或对 DNA 进行液体闪烁计数。

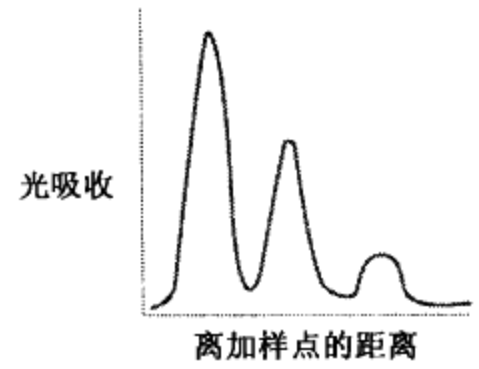


图 5.9　光密度计。光密度扫描图下方是相应的放射自显影图。注意 3 个扫描峰的面积与放射自显影图中相应条带的黑色程度成正比。

磷屏成像

磷屏成像在某些方面胜过常规的放射性自显影，最主要的是它能更精确地定量放射性，它对放射性的感应比 X 线胶片线性化得多。常规放射性自显影的 50 000dpm（**每分钟蜕变量** disintegration per minute）条带可能看上去并不比 10 000dpm 的条带黑，因为感光胶片在 10 000dpm 已经饱和。而**磷屏成像仪**（phosphorimager）是在电子水平对放射性进行检测与分析，所以 10 000dpm 与 50 000dpm 的区别是很明显的。该技术的工作原理是：从放射性样品如一块带有与标记性探针杂交的 RNA 条带的印迹膜开始，然后将其放到一个吸收 β 电子的磷屏成像板上。β 电子激发板上的分子，这些分子就一直保持激发状态直到磷屏成像仪用光束扫描成像板。此时，成像板捕获的 β 电子能被释放出来并被微机

化检测仪监控。计算机将检测到的能量转换成如图5.10所示的图像模式。这是模拟的彩色图像，由黄到黑等不同颜色分别代表不同程度的放射性从最低（黄色）到最高（黑色）。

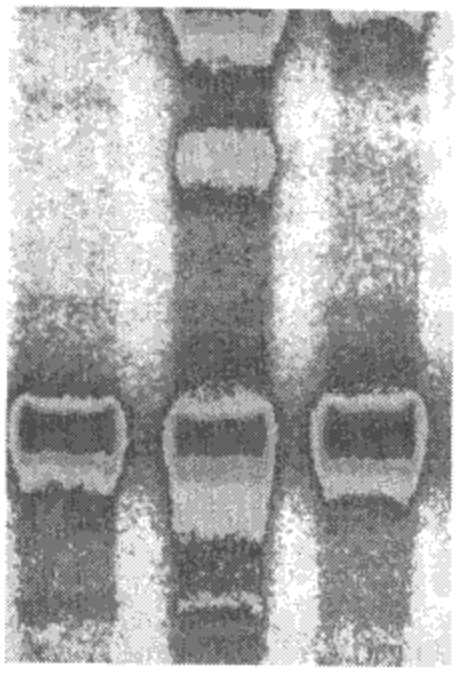


图5.10　RNA印迹膜的假彩色磷屏成像扫描。RNA印迹膜经放射性同位素探针杂交后，洗去未杂交的探针，将印迹膜置于磷屏成像仪的扫描板上。扫描板收集结合在RNA上的放射性探针发出的β电子的能量，当激光扫描时放出能量。计算机再将这些能量转换成图像，放射强度按照以下颜色递增：黄色（最低）＜紫色＜紫红色＜淡蓝＜绿色＜深蓝＜黑色（最强）。（*Source*：© Jay Freis/Image Bank/Getty.）

液体闪烁计数器

液体闪烁计数（liquid scintillation counting）利用样品放射性发出的射线产生的一种能被光电倍增管检测到的可见光光子。为此，将放射性样品（如从凝胶上切下的一个条带）放到含有**闪烁液**（scintillation fluid）的小瓶中，闪烁液含有化合物**氟石**（fluor），被放射线轰击后可发出荧光，从而将非可见的放射线转换成可见光。液体闪烁计数仪将小瓶降到带有光电倍增管的暗室中，光电倍增管检测由放射线激发荧光物所发出的光，仪器计数光的突发或**闪烁**（scintillation），单位是每分钟的次数（counts per minute，**cpm**）。但这与每分钟的蜕变不一样，因为液体闪烁计数不是100%有效的。分子生物学家常用的放射性同位素是^{32}P，由这种同位素发出的β电子能量极高甚至不需要荧光就能产生光子，可用液体闪烁计数器直接计数，但比用闪烁液时的效率要低些。

小结　分子生物学实验中的微量物质通常用标记示踪物来检测。如果示踪物是放射性的，可通过X线胶片放射自显影、磷屏成像或液体闪烁计数法检测。

非放射性示踪剂

正如我们在本节前面部分所指出的，放射性示踪的最大优点是它的灵敏性，而现在非放射性示踪物的灵敏度已经可与其放射性的前辈竞争了。这是

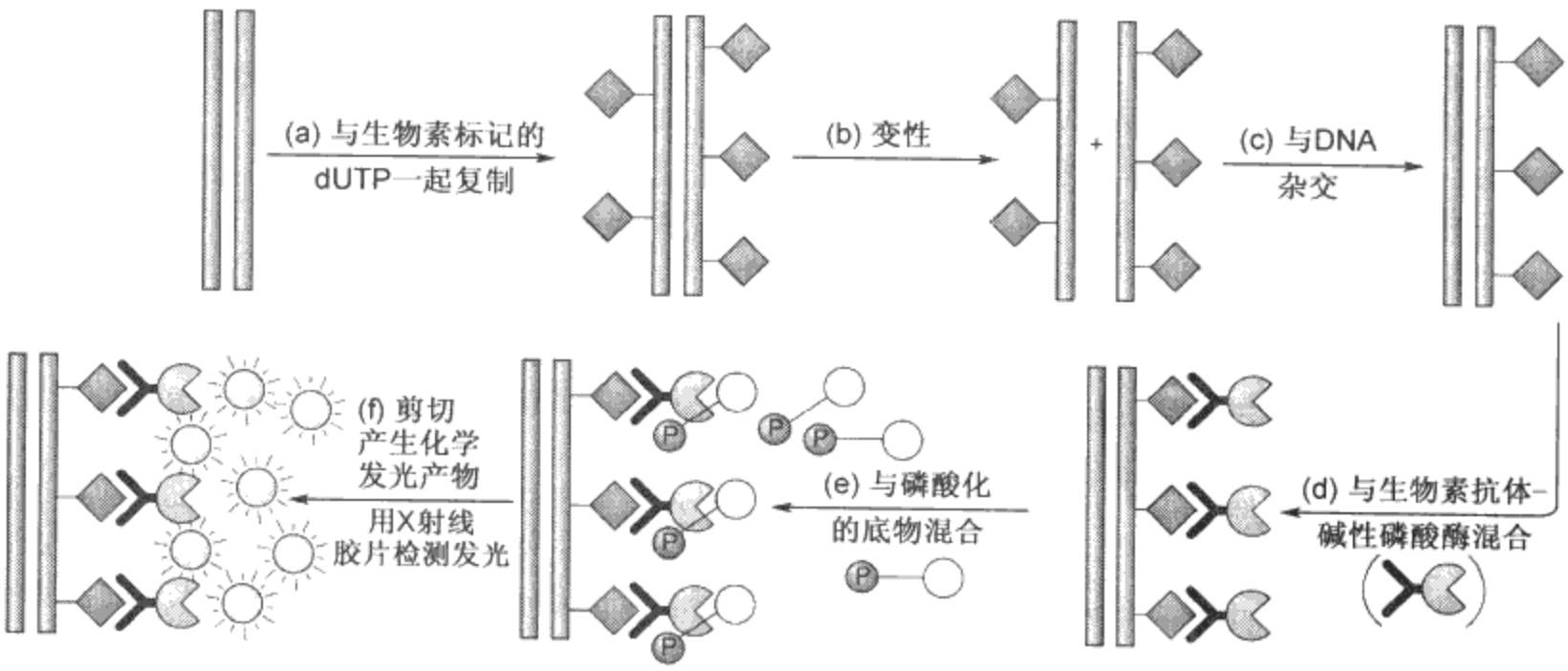


图5.11　用非放射性探针检测核酸。这类技术通常是间接的，通过与标记的探针杂交检测目的核酸，而对探针的检测又依赖其产生有色或发光物质的能力。在本例中实施了以下操作步骤。（a）在带有生物素标签（菱形）的dUTP存在时复制探针DNA，合成生物素标记的探针DNA；（b）探针变性；（c）将探针与待检测的DNA（粉红色）杂交；（d）将杂交产物与含有生物素抗体和碱性磷酸酶（缺口圆与叉）的双功能试剂混合，生物素抗体特异地与探针上的生物素紧密结合；（e）加入一种磷酸化的复合物，一旦其磷酸基被去除就转变为化学发光物质；（f）附着在探针上的碱性磷酸酶将磷酸基团从底物上切除，使之成为化学发光物质（发射光）。化学发光物质发出的光可用X射线胶片检测。

一个显著的优势，因为放射性同位素具有潜在的危害健康的问题，使用时必须非常小心。而且放射性示踪物产生放射性废物，处置这些废物日益困难且很昂贵。非放射性示踪物怎样与放射性示踪物的灵敏度竞争呢？答案是利用酶的倍增效应。将酶偶联到用于检测目的分子的探针上，酶会产生许多产物分子从而将检测信号放大。若酶的产物是化学发光的（chemiluminescent）（如萤火虫尾部的发光），则效果会更好，因为每个分子都发出许多光子从而将信号再次扩大。图 5.11 是一种化学发光示踪物的原理，发出的光可以通过 X 线胶片放射自显影或磷屏成像检测。

为避免磷屏成像或 X 线胶片的花费，也可以使用能发生颜色变化而不是形成化学发光物的酶底物。这些**发色底物**（chromogenic substrate）产生与酶的位置相对应的有色条带，即与目的分子的位置相对应。颜色的亮度与目的分子的量直接相关，所以也是一种定量方法。

小结 现在已经有非常灵敏的非放射性标记示踪物了。利用化学发光示踪物可以像放射性那样通过放射自显影或磷屏成像来检测，而产生有色产物的示踪物可通过直接观察出现的有色斑点进行检测。

5.3 利用核酸杂交

杂交现象，即一条核酸单链和另一条与之互补碱基序列的单链形成双螺旋的能力，是现代分子生物学的主干技术之一。我们在第 4 章接触过噬菌斑及菌落杂交，这里进一步介绍几个杂交技术的例子。

Southern 印迹：鉴定特异性 DNA 片段

许多真核基因都是亲缘关系较近的基因家族的一部分。那么，怎样确定某个特定基因家族的成员数目呢？如果已经克隆到了这个基因家族的一个基因，甚至是“部分”cDNA，那么就能估算出这个数目。

首先用限制酶切割从生物体提取的基因组 DNA。最好用 *Eco*RI 或者 *Hind*III 等识别 6bp 的限制酶酶切位点。这些酶会产生数千个平均大小约 4000bp 的基因组 DNA 片段。然后，在琼脂糖凝胶上电泳分离这些片段（图 5.12）。如果通过染色显示，其结果将是数千条带组成的模糊条纹，无法区分（简洁起见，图 5.12 只显示了几个条带）。最后，用标记的探针与这些条带杂交，就可以看到其中有多少含有目的基因的编码序列。当然，在此之前要先把这些条带转移到一个方便杂交的介质上。

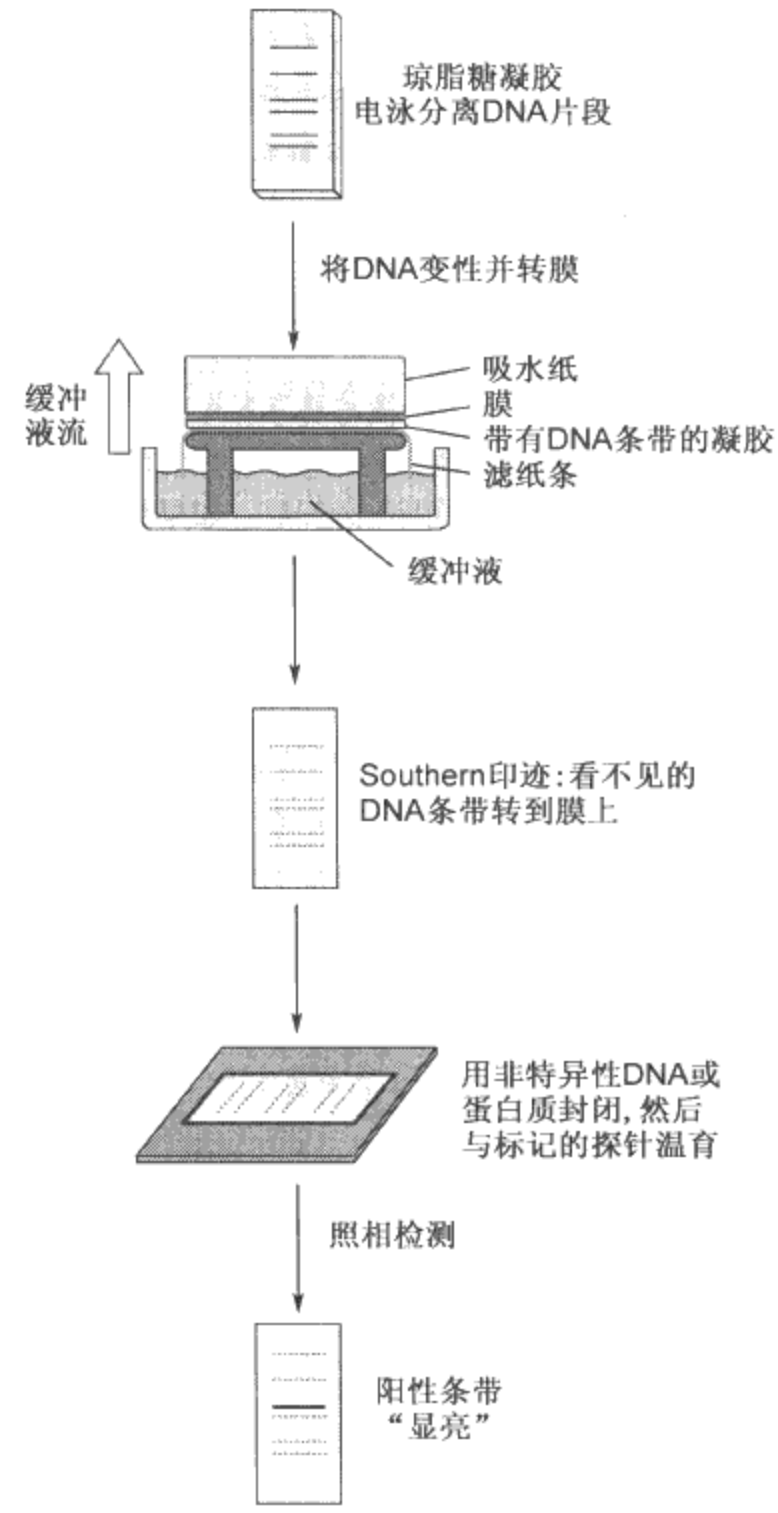


图 5.12 Southern 印迹。首先，在琼脂糖凝胶上对 DNA 片段进行电泳。接着，对 DNA 进行碱变性，将单链 DNA 从凝胶上转移到硝酸纤维素膜或类似材料上。可用两种方法完成这一步：扩散法，如图中所示，DNA 随着缓冲液通过凝胶；另一种是电泳法（未显示）。接下来，将标记的探针与印迹膜杂交，通过放射自显影或磷屏成像检测被标记的条带。

Edward Southern 是这一技术的创始人。如图 5.12 所描述的那样，他将琼脂糖凝胶上的 DNA 片段通过扩散法转移或者印迹到硝酸纤维膜上，因此这一过程被叫做 **Southern 印迹**（Southern blot）。现在多采用电转移将 DNA 条带从凝胶转移到膜上。在转膜印迹之前，将 DNA 片段碱变性产生单链 DNA 以便与硝酸纤维膜结合，形成 Southern 印迹。现在已有比硝酸纤维素膜更好的介质可供使用，尼

龙膜介质比硝酸纤维膜的柔韧性更强。接下来，在标记的 DNA 前体存在条件下，加入 DNA 聚合酶对克隆的 DNA 进行标记，再将标记的探针变性并与 Southern 印迹膜杂交。当探针遇到与之互补的 DNA 序列时，就杂交形成被标记的条带，与之相对应的 DNA 片段含有目的基因。最后，通过 X 射线胶片放射性自显影或磷屏成像就可看见被标记的条带。

如果只看到一个条带，解释起来相对简单，可能只有一个基因含有与 cDNA 探针互补的序列。一个基因（如组蛋白基因或核糖体 RNA 基因）可能多次重复串联，在基因的每个拷贝中只有一个限制酶位点，这会得到一条很黑的条带。如果看到的是多个条带，可能有多个基因与探针相匹配，但是很难准确判断出到底有多少个。当一个基因中含有一个以上酶切位点时也可以形成多个条带，使用短探针可减小这一问题，如用 100～200bp 的 cDNA 限制酶酶切片段作探针。一个平均约每 4000bp 切割一次的限制酶，不会在与小探针杂交基因的 100～200bp 区域内切割。如果用短探针杂交后仍得到大量条带，这可能代表的是一个基因家族，其家族成员的序列与探针杂交的区域相似或者完全相同。

小结 标记的 DNA（或 RNA）探针可用于与 Southern 印迹膜上序列相同或相近 DNA 的杂交。与短探针杂交的条带数可以估计出在一个生物中紧密相关基因的数目。

DNA 指纹技术和 DNA 分型

Southern 印迹技术不仅仅是一种研究工具，而且还广泛应用于法医实验室以鉴别在犯罪现场留下的血迹或其他含 DNA 物质的个体。这种 **DNA 分型**（DNA typing）源自于 1985 年 Alec Jeffreys 及同事的一个发现。当时他们正在研究人类血液中 α-球蛋白基因的一个 DNA 片段，发现其中有一段含数个碱基重复的序列，这种重复的 DNA 叫做**小卫星**（minisatellite）。更有趣的是，他们还在人类基因组的其他位置发现了类似的小卫星序列，而且同样有数次重复。这一看似简单的发现产生了一个影响深远的结果，因为不同个体基本序列的重复方式不一样。实际上这种差异有足够大，使任意两个个体具有完全相同重复方式的概率是随机的，意味着这些重复方式就像指纹一样，因此把它们叫做 **DNA 指纹**（DNA fingerprint）。

DNA 指纹实际上就是一个 Southern 印迹。为获得 DNA 指纹，研究者首先用限制酶（如 *Hae*III）切割所研究的 DNA，选用 *Hae*III 是因为在 Jeffreys 发现的这种重复序列中不含 *Hae*III 酶切位点。如图 5.13a 所示，这表明 *Hae*III 在小卫星区域的两边而不会在其内部切割。这种情况下，DNA 有三组重复区域，分别含有 4、3、2 个重复，酶切后会出现包含这些重复序列的 3 个大小不等的片段。将这些片段电泳、变性、印迹，然后用一个标记的小卫星 DNA 与该印迹杂交，杂交后用 X 射线胶片自显影或磷屏成像检测标记的条带。本例中有 3 个条带被标记，所以在胶片上会出现 3 个黑色条带（图 5.13d）。

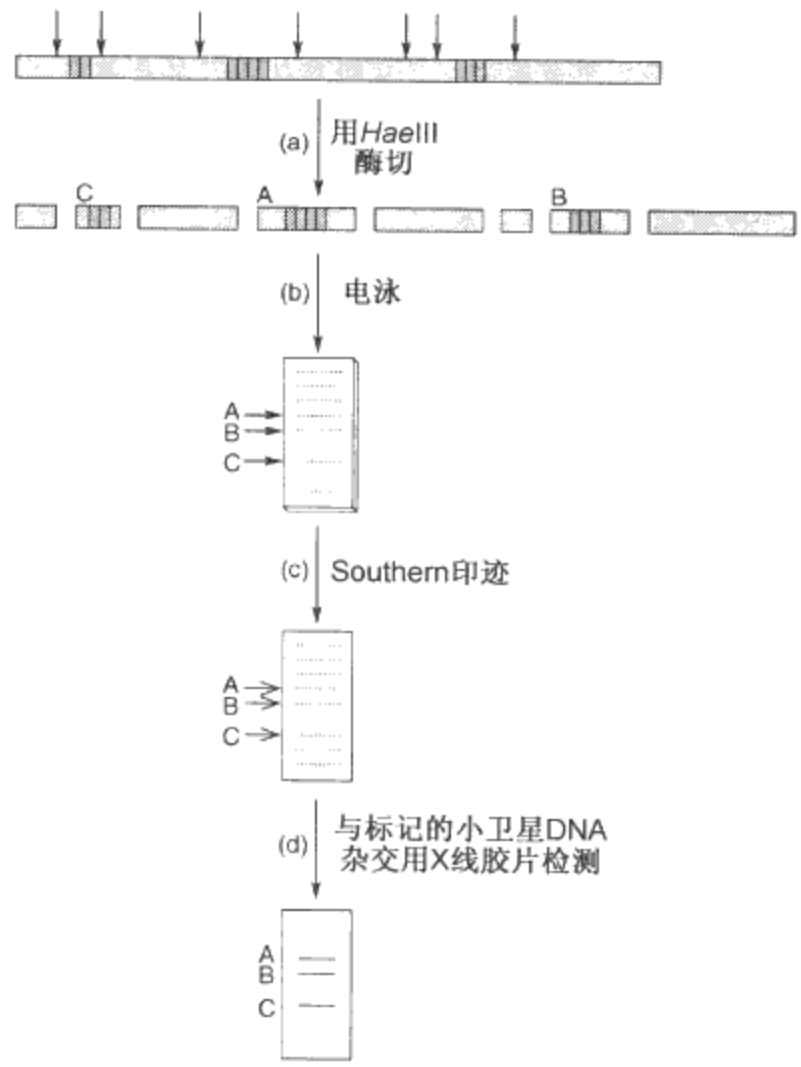


图 5.13 DNA 指纹技术。（a）首先，用限制酶切割 DNA。本例中 *Hae*III 酶在 7 个位点切割 DNA（箭头所示），产生 8 个片段，仅 3 个片段（按大小标记为 A、B、C）含有微卫星 DNA，用蓝色盒表示。其他片段含不相关 DNA 序列。（b）电泳（a）中的片段，使其按大小分离开。所有 8 个片段都在凝胶中，但是看不见，用虚线表示。（c）将 DNA 片段变性并进行 Southern 印迹。（d）将Southern 印迹膜上的 DNA 片段与带有若干个小卫星拷贝的放射性 DNA 杂交。探针与带有小卫星的 3 个片段结合，不与其他几个片段结合。最后，用 X 线胶片检测 3 个标记的条带。

真实的动物基因组要比这个例子中的 DNA 片段复杂得多，因此含有与探针反应的微卫星序列的片段会远远多于 3 个。图 5.14 是几个无关人员以及一对同卵双生子的 DNA 指纹。正如我们前面所说的，图中任意两个个体之间重复片段的差异非常大，除非他们是同卵双生子，这种复杂性使 DNA 指纹成为一种非常有效的鉴定技术。

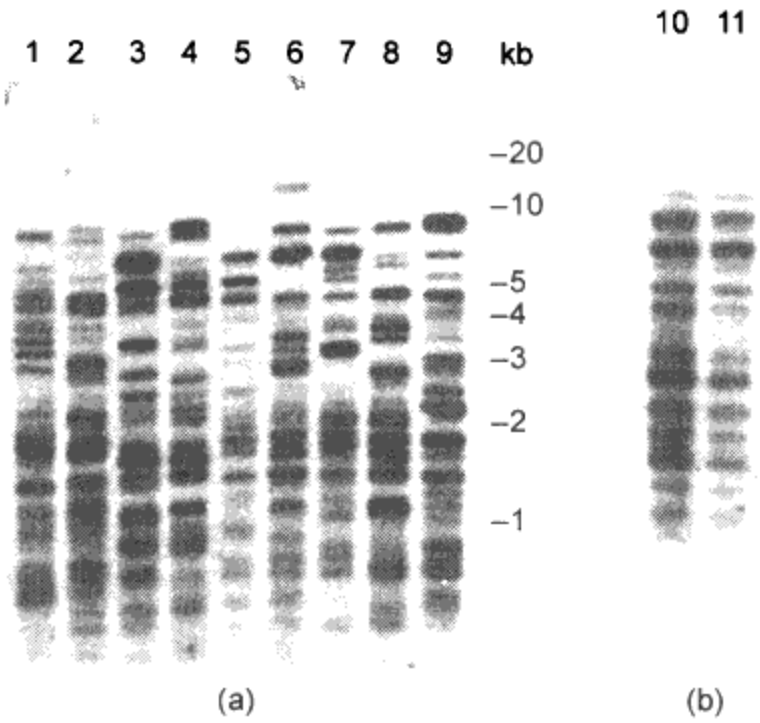


图 5.14　DNA 指纹。（a）9 个泳道中的 DNA 样品来自 9 个不相关的白人受试者。注意没有两个完全一样的图谱，特别是在上端部分。（b）两条泳道中的 DNA 来自同卵双生者，所以图谱相同（虽然 10 道的 DNA 重量比 11 道的多）。（*Source*: G. Vassart et al., A sequence in M13 phage detects hypervariable minisatellites in human and animal DNA. *Science* 235 (6 Feb 1987) p. 683, f. 1. © AAAS.）

DNA 指纹技术和 DNA 分型在法医学中的应用

DNA 指纹的一个宝贵特征是，虽然几乎所有个体都有不同的图谱（pattern），但是部分图谱（几组条带）是按照孟德尔定律遗传的。所以，DNA 指纹可用于建立亲子关系。英格兰的一桩移民案很好地诠释了这一技术的威力。有一个出生在英格兰的加纳男孩返回加纳与他的父亲一起生活，当他想返回英国与母亲一起生活时，英国当局怀疑他是那位妇人的儿子还是侄子。来自血型基因的信息十分模糊，但是这个男孩的 DNA 指纹证实了他们之间的确是母子关系。

除了亲子鉴定外，DNA 指纹技术还可用于罪犯鉴定，因为理论上每个人的 DNA 指纹都是独特的，就像传统意义上的指纹。因此如果罪犯把他的细胞（如血液、精液、头发等）留在犯罪现场，通过这些细胞的 DNA 就可以确认他。然而，正如图 5.14 中所示，DNA 指纹含有数百个条带，非常复杂，有些条带混在一起难以辨别。

为解决这一问题，法医学家开发了新的探针，这些探针与单一的、在个体中有很大变异的 DNA 位点杂交，而不是像标准 DNA 指纹那样与所有的 DNA 位点都杂交。每个探针产生非常简单的样式，只有一个或几个条带。这就是**限制性片段长度多态性**（restriction fragment length polymorphyism, RFLP）的一个例子，此方法将在第 24 章详细介绍。能够产生 RFLP 是由于指定位点的限制性片段的大小在个体间有差异。当然，单个探针本身不能像具有众多条带的完整 DNA 指纹那样作为一个有力的鉴定工具，但是用 4 或 5 个探针的组合就能提供足够多的不同条带用于鉴定。有时我们仍把这种分析方法叫做 DNA 指纹，但实际上更具体的术语是 DNA 分型（DNA typing）。

有一个早期的关于 DNA 分型的戏剧性故事。说是有一个人趁一对夫妇熟睡时将他们杀死在一辆敞篷卡车中，40min 后他又返回案发地强奸了那名妇女。这一行为不仅加重了犯罪，也给法医学家提供了证明罪犯的方法。法医学家从他留下的精细胞中提取了 DNA 并进行 DNA 分型，结果表明其 DNA 分型图谱与嫌疑人的完全匹配，这一证据帮助陪审团裁定了被告有罪。图 5.15 是用 DNA 分型鉴定另一个强奸嫌疑人的例子，嫌疑人的 DNA 分型图谱很明显与遗留精液中的 DNA 相匹配。这是单一探针杂交产生的结果，其他探针同样也提供了与精子 DNA 匹配的样式。

DNA 分型的另一个优点是其高度的灵敏性，只需要几滴血液或精液就足以完成一次检测。但是有时法医学家得处理甚至更少的样品，如被受害者扯下的一根头发。虽然这根头发本身不足以做 DNA 分型，但是如果它带有毛囊细胞就有用了，从这些细胞选择的 DNA 片段可以通过 PCR 扩增，然后进行分型。

尽管 DNA 分型具有精确性，但是在法庭上还是受到了有效地挑战，最著名的就是 1995 年洛杉矶的 Simpson 案件。被告辩护律师把焦点集中在两个有关 DNA 分型的问题上。首先是由于 DNA 分型的灵敏度非常高，操作时必须十分小心才能得到有意义的结果；二是对得到的数据进行分析所使用的统计学方法一直有争议。第二个问题的核心是围绕应用乘法定则（product rule）确定 DNA 分型结果是否能唯一地鉴定嫌疑人的问题。例如，用一个给定探针从一般人群的 100 个人中探测某个人的给定等位基因（在这个例子中是一系列条带）。该探针与指定的某个人相匹配的概率是 1%（即 10^{-2}），如果我们用 5 个探针，并且 5 个等位基因都与嫌疑人匹配，那么这种匹配的概率就是每个探针匹配概率的乘积，即 $(10^{-2})^5$ 或 10^{-10}。因为现在地球人口不到 10^{10}（100 亿）人，这就意味着 DNA 分型从统计学上排除了嫌疑人之外的其他人。考虑到一些少数民族与

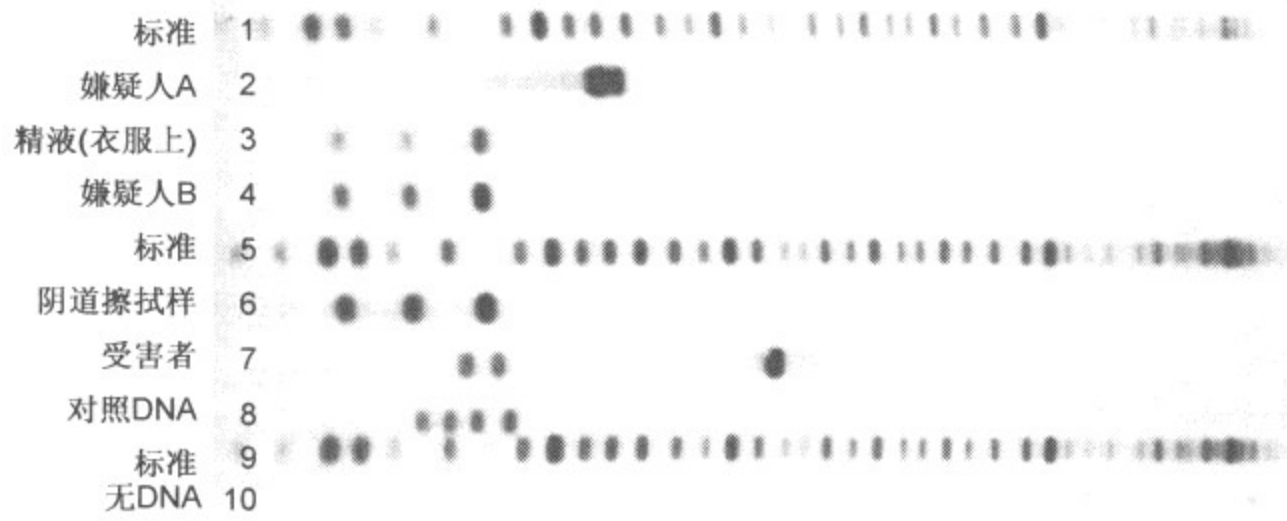


图 5.15　利用 DNA 分型辅助鉴定强奸犯。两名嫌疑人被指控袭击和强奸一位年轻妇女，对嫌疑人和受害者的多个 DNA 样本进行了分析。泳道 1、5 和 9 是已知的标准 DNA，泳道 2 是嫌疑人 A 的血细胞 DNA，泳道 3 是受害人衣服上残留的精液样本 DNA，泳道 4 是嫌疑人 B 的血细胞 DNA，泳道 6 是受害人阴道擦拭样本的 DNA（也有微量受害者自身的 DNA），泳道 7 是受害者的血细胞 DNA，泳道 8 是 DNA 阳性对照。泳道 10 是无 DNA 的阴性对照。以这些证据为基础，嫌疑人 B 被判有罪。注意泳道 4 的 DNA 片段与泳道 3 受害者衣服上的精液样本和泳道 6 阴道拭擦样本的 DNA 一致。

某些探针可能有更高的匹配概率，起诉人已经用了比较保守的估计。然而，大于百万分之一的匹配率仍会不断出现，因此在法庭上具有极强的说服力。当然，DNA 分型不只用于鉴定罪犯，它同样可以用于排除嫌疑人（图 5.15 中嫌疑人 A）。

小结　现代 DNA 分型技术使用一组 DNA 探针鉴定动物个体包括人的可变位点。作为一种取证工具它可用于亲子鉴定、罪犯鉴定及排除无辜人员的嫌疑。

原位杂交：基因在染色体中的定位　本章已经介绍了利用探针鉴定 Southern 印迹膜上的条带是否含有目的基因。标记的探针还可与染色体杂交用于鉴定哪个染色体含有目的基因。**原位杂交**（*in situ* hybridization）的原理是将细胞中的染色体展开，将 DNA 部分变性以产生能与探针杂交的单链区域。染色体被染色、杂交后，用 X 线胶片检测展开的染色体上被标记的基因。染色的目的是为了能看清并辨认出不同的染色体，而感光底片的黑点可以定位标记的探针从而显示与此探针杂交的基因。

也可用其他方法标记探针。图 5.16 是二硝基酚标记的 DNA 探针将肌糖原磷酸化酶基因定位在人 11 号染色体上，这种二硝基酚可用荧光抗体检测。染色体被碘化丙啶复染色发出红色荧光，在这种背景下 11 号染色体上探针抗体的黄色荧光很容易看到。这种技术也叫做**荧光原位杂交**（fluorescence *in situ* hybridization，FISH）。

小结　将标记的探针与整个染色体杂交可定位基因或其他特定的 DNA 序列，这种技术叫做原位杂交。如果探针是荧光标记的，那么这种技术就叫做荧光原位杂交 FISH。

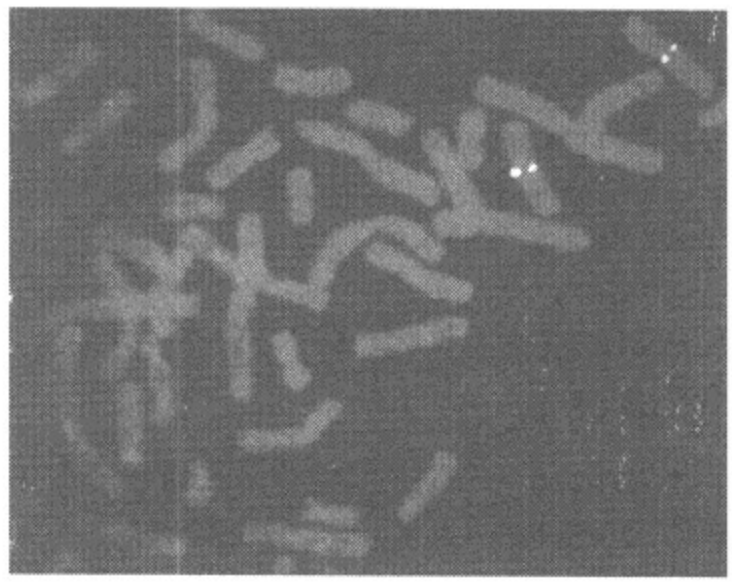


图 5.16　利用染色体荧光探针原位杂交寻找基因。人肌糖元磷酸化酶基因的特异性 DNA 探针与二硝基酚偶联。将展开的人染色体部分变性暴露出单链 DNA 区域以便探针杂交。间接检测 DNP 标记探针杂交位点的方法如下：兔抗 DNP 的抗体与探针上的 DNP 结合，联有异硫氰酸荧光素（FITC）的羊抗兔抗体再与兔抗体结合，FITC 发黄色荧光。探针与染色体的杂交位点显示亮的黄色荧光斑点，背景是被荧光燃料碘化丙啶染色的红色染色体。分析表明肌糖元磷酸化酶基因位于 11 号染色体上。（*Source*：Courtesy Dr. David Ward，*Science* 247（5 Jan 1990）cover. © AAAS.）

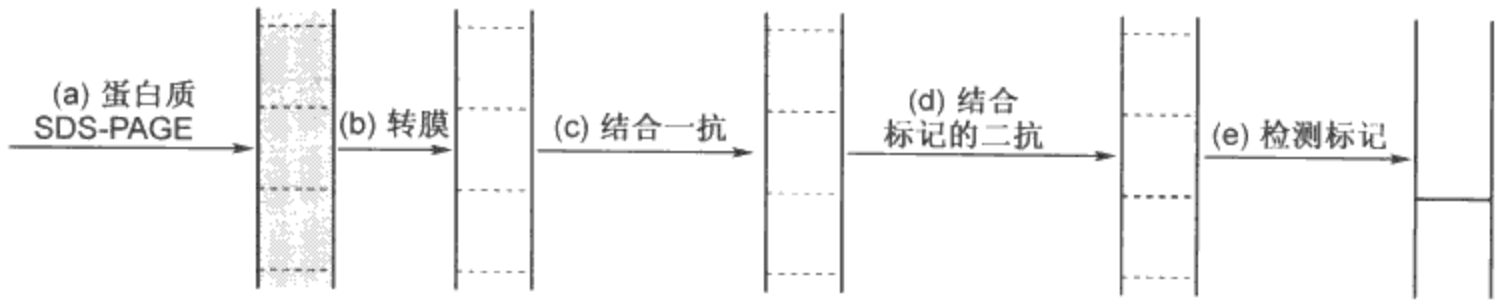


图 5.17　免疫印迹（Western 印迹）。（a）从 SDS-PAGE 分离蛋白质混合物开始；（b）分离开的蛋白质（虚线表示）被印记到膜上；（c）用原初抗体探测膜上的目的蛋白，这里抗体与其中一个蛋白质发生了反应（红色），但是目前还不能检测该反应；（d）用标记的第二抗体（或蛋白质 A）检测一抗进而检测目的蛋白，这里二抗结合到一抗上，条带的颜色由红色变为紫色，但是反应还是不能检测；（e）最后，检测标记的条带，如果是放射性标记的用 X 线胶片或磷屏成像仪，如果是非放射性标记的可用图 5.11 所示的方法检测。

Western 印迹

免疫印迹（immunoblot）（也叫做 Western 印迹，沿用了 Southern 的命名系统）采用与 Southern 印迹相同的实验模式：将分子电泳后印迹到膜上再做鉴定。但是免疫印迹是对蛋白质电泳而不是核酸。我们已经看到 Southern 印迹膜上的 DNA 是通过与标记的寡核苷酸或多核苷酸探针杂交后鉴定的，而杂交作用只适用于核酸，如何对印迹的蛋白质做鉴定呢？需要采用对特定蛋白质的专一性抗体（或抗血清）。这个抗体与印迹膜上的目的蛋白结合，然后用标记的第二抗体（如在免疫球蛋白 IgG 中识别所有兔抗体的羊抗体），或标记的 IgG 结合蛋白（如蛋白质 A）与已经结合上的一抗结合再标记带有目的蛋白的条带（根据抗体是免疫系统的产物而产生了“免疫印迹”这一术语）。免疫印迹可以告诉我们混合物中是否存在某种特殊蛋白，也能给出该蛋白质粗略的定量。

为什么要用第二抗体或蛋白 A，而不直接用标记的原初抗体呢？主要原因是这需要一一标记不同的抗体用于一系列免疫印迹膜，而用不标记的原初抗体更简单和便宜，可以购买与任何一抗都能结合并检测的标记好的二抗或蛋白 A。图 5.17 给出了对特定蛋白质制备和探测免疫印迹膜的过程。

DNA 测序

1975 年，Frederick Sanger 及同事，以及 Alan Maxam 和 Walter Gilbert 研究出两种不同的方法测定克隆的 DNA 片段中准确的碱基序列。这一激动人心的重大突破使分子生物学发生了革命，Sanger 和 Gilbert 也因此获得 1980 年诺贝尔化学奖。这两种方法使分子生物学家能够确定数千种基因的碱基顺序，以及许多完整基因组（包括人类基因组的几乎 30 亿个碱基对）。现代 DNA 测序技术源于 Sanger 法，这里对这种方法做一介绍。

Sanger 链末端终止测序法（Sanger chain-termination sequencing method）　用 Sanger 最初的方法测序一段 DNA 序列的原理见图 5.18。在应用上，该过程不需再手动操作了，在下一节将介绍该技术是如何自动操作的。最初的测序方法首先需要将 DNA 克隆到载体中，如 M13 噬菌体载体或者噬菌粒，产生单链 DNA。现在可以从双链 DNA 开始了，通过简单的加热产生单链 DNA 用于测序。然后将一段长约 20 个碱基的寡核苷酸引物与单链 DNA 杂交。引物与载体多克隆位点相邻的一段序列杂交，并且其 3′端朝向多克隆位点处的插入片段。

用 DNA 聚合酶的 Klenow 片段（第 20 章）延伸引物，合成与插入片段互补的 DNA。Sanger 法的技术要点在于 DNA 合成反应分别在 4 个试管中独立进行，且每管各有一种不同的链终止剂。链终止剂是一种**双脱氧核苷酸**（dideoxy nucleotide），如**双脱氧 ATP**（ddATP）。这种终止剂不仅像正常的 DNA 前体物那样在 2′位脱氧，它的 3′位也是脱氧的。终止剂由于缺乏必需的 3′-羟基，不能形成磷酸二酯键，因此称为链终止剂。一旦双脱氧核苷酸掺入延长中的 DNA 链上，DNA 合成就会终止。

双脱氧核苷酸本身根本不允许 DNA 合成，所以必须加入过量的脱氧核苷酸和足够的双脱氧核苷酸以便随机终止 DNA 链的延伸。DNA 生长的随机停止意味着有的 DNA 链终止的早些，有的则晚些。每个试管中各有一种双脱氧核苷酸：试管 1 是 ddATP，因此终止全部发生在 A 处。试管 2 是 ddCTP，终止全部发生在 C 处。其他两个试管依此类推。同时，每个试管中都有放射性 dATP，这样 DNA 产物就具有放射性。

结果每个管中都产生了一系列长度不等的片段。在试管 1 中全部片段都以 A 结尾；试管 2 中全部以 C 结尾；试管 3 中以 G 结尾；试管 4 中以 T 结尾。

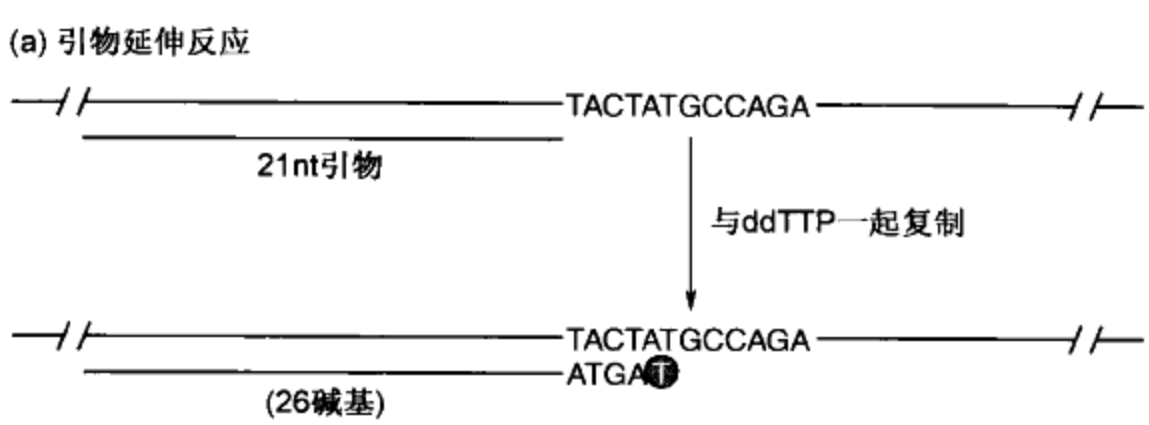


图 5.18　DNA 测序的 Sanger 双脱氧法。（a）引物延伸（复制）反应。本例中，用 21nt 的引物与待测序的单链 DNA 杂交，然后与 DNA 聚合酶的 Klenow 片段和 dNTP 底物混合进行复制。每管中含有一种双脱氧 NTP，会在一定的碱基之后终止复制，这里用的是 ddTTP，使反应终止在的第二个 dTTP 上。（b）4 个反应的产物。顶部显示的是模板链，下面是各种产物。每个产物都从 21nt 的引物开始，在 3′端延伸一个或多个核苷酸，最后一个核苷酸总是终止反应的双脱氧核苷酸（带圈的）。每个产物的长度在片段左端的括号中给出，22～33 个碱基。（c）产物的电泳。4 个反应的产物分别加到高分辨率凝胶的平行泳道中，按大小电泳分离开。从底部开始，找出最短的片段（A 泳道的 22nt），次短的片段（T 泳道的 23nt），依次延长就可以读出 DNA 产物的序列。当然这是模板链的互补序列。

接着，将这 4 组反应产物在变性条件下在高浓度聚丙烯酰胺凝胶平行的泳道内电泳。由于是在变性条件下电泳，故所有 DNA 都是单链的。最后进行放射自显影显现 DNA 片段，在 X 线胶片上以平行条带出现。

图 5.18c 是一张测序胶片的示意图。序列阅读从胶片底部的第一个条带开始读起。在这里，第一个条带在 A 泳道，所以该短片段是以 A 结尾。现在移到下一个稍长的片段，即胶片的向上一个片段。凝胶电泳具有极高的分辨率，至少在片段合成到非

常长之前，仅一个碱基之差的片段都能被分离开。下一个片段比第一个多一个碱基，出现在 T 泳道，所以一定是以 T 结尾。因此，我们就得到了序列 AT。按照这种方法接着读，从下向上就会获得一个完整的序列如图右侧所示。刚开始时，读到的只是载体多克隆位点的部分序列，但是不会太长 DNA 链就会延伸至插入序列以及未知区域。有经验的测序者从一张胶片中可连续读出几百个碱基序列。

图 5.19 是一张典型的测序胶片。最短的片段（在最底部）位于 C 泳道中，之后连续 6 个条带均位于 A 泳道。所以该序列以 CAAAAAA 开始。很容易从这张胶片上读出更多的碱基顺序，你可以自己试一下。

图 5.19　典型的测序胶片。序列从 CAAAAAACGG 开始，由下到上可以读出整个序列。(*Source*: Courtesy Life Technologies, Inc., Gaithersburg, MD.)

自动化 DNA 测序　上述这种“手动”测序技术十分有效，但是仍然相对较慢。如果要测定非常大的 DNA，如人类基因组的 30 亿个碱基对，那么就需要快速自动化的测序方法。事实上 DNA 自动化测序技术已经使用了多年。图 5.20a 是一种自动化测序技术，其工作原理仍是基于 Sanger 的链终止法。所使用的双脱氧核苷酸，除用不同的荧光分子标记外，与手动法一样。所用的引物，或更通俗地说，4 组反应中的双脱氧核苷酸分别用不同的荧光分子标记，每个试管的产物在光激发下会发出不同颜色的荧光。

在延伸反应和链终止完成后，将全部 4 组反应物混合，在一个短的凝胶细柱内进行同一泳道电泳（图 5.20b）。凝胶底部附近是一个分析器，反应混合液通过时它能用一个激光束激发带荧光素的寡核苷酸，然后用电子装置检测每条寡核苷酸发出的荧光颜色，同时将信息传输到电脑中，通过设定的程序将颜色信息转化成碱基序列。如果电脑“看到”蓝色，就意味着这个寡核苷酸来自双脱氧 C 反应，因此以 C 结尾（实际是 ddC）。绿色指示 A、橙色 G、红色 T。计算机打印出所检测的每个荧光条带的图谱，用不同颜色表示碱基（如图 5.20c），并将碱基序列储存起来备用。

现在，自动测序仪（automated sequencer 或 sequenator）可以很简单地打印出序列，或直接将其输入计算机进行分析。大型基因组计划使用多台 96 道甚至 384 道测序仪同时运行以获得数百万甚至上亿个碱基序列（见第 24 章）。一台 384 道测序仪运行 3h 可产生 200 000nt 的序列。

小结　Sanger DNA 测序法采用双脱氧核苷酸终止 DNA 合成，产生一系列 DNA 片段，其大小可通过电泳检测。因为已经知道每个反应由哪种双脱氧核苷酸终止，因此每个片段的最后一个碱基都是已知的。将这些片段按照大小排列，每个片段比下一个片段多一个碱基（是已知的），就可以知道这段 DNA 的碱基顺序。

限制性图谱

在对一个大的 DNA 片段测序之前，一般要初步绘制一张图谱，对 DNA 分子上的标记进行定位。这些标记不是基因而是一小段 DNA 序列，如限制酶的酶切位点。这种基于物理性质的图谱很自然地叫做**物理图谱**（physical map）（如果仅以限制酶酶切位点作为标记，则叫做**限制性图谱** restriction map）。

为了介绍限制性图谱的做法，让我们看一下图 5.21 的一个简单例子。以一个 1.6kb（1600bp）的 *Hind* III 酶切片段（图 5.21a）开始。用另一个限制酶 *Bam*HI 酶切该片段后得到两条长度分别为 1.2kb 和 0.4kb 的片段，其大小通过电泳测定如图 5.21a。切出的大小表明，*Bam*HI 从 1.6kb-*Hind* III 片段的一端切掉 0.4kb，另一端切掉 1.2kb。

现在，假设把 1.6kb *Hind* III 片段克隆到一个假定质粒载体的 *Hind* III 位点，如图 5.21b 所示。由于不是定向克隆，该片段可能以两种方向插入载体：*Bam*HI 酶切位点在右侧（图 5.21 左边的图），或 *Bam*HI 酶切位点在左侧（图 5.21 右边的图）。那么，对于给定的克隆如何确定片段的插入方向呢？

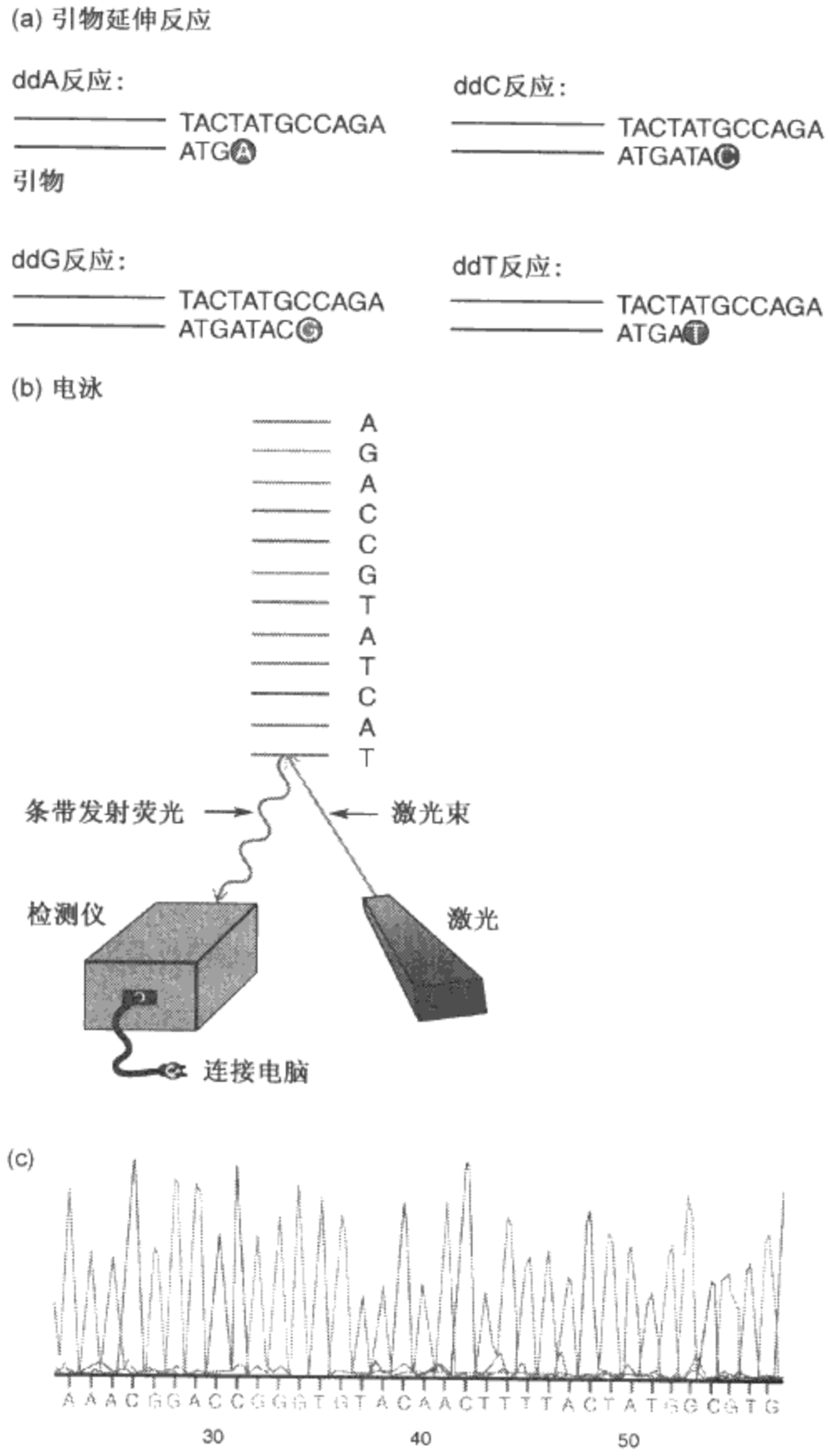


图 5.20　自动化的 DNA 测序。（a）引物延伸反应与手动测序一样，只是在每个反应中引物用不同的荧光分子标记发出不同颜色的光。每个反应只显示了一个产物，但实际上就像手工测序一样所有可能的产物都有。（b）条带的电泳和检测。通过电泳分离各种大小的引物延伸产物。根据引起终止反应的碱基类型的颜色表示条带（即绿色是以 ddA 结束的，蓝色是以 ddC 结束的等）。激光扫描仪扫描每个条带时，激发其荧光标签，检测仪分析所激发光的颜色。这些信息被转化成碱基序列并存储在计算机中。（c）DNA 自动化测序结果的输出。每个颜色的峰图表示条带通过激光束时的荧光强度。图中所选峰的颜色、（b）中条带的颜色、（a）中标签的颜色是为了方便，可能和实际不一致。

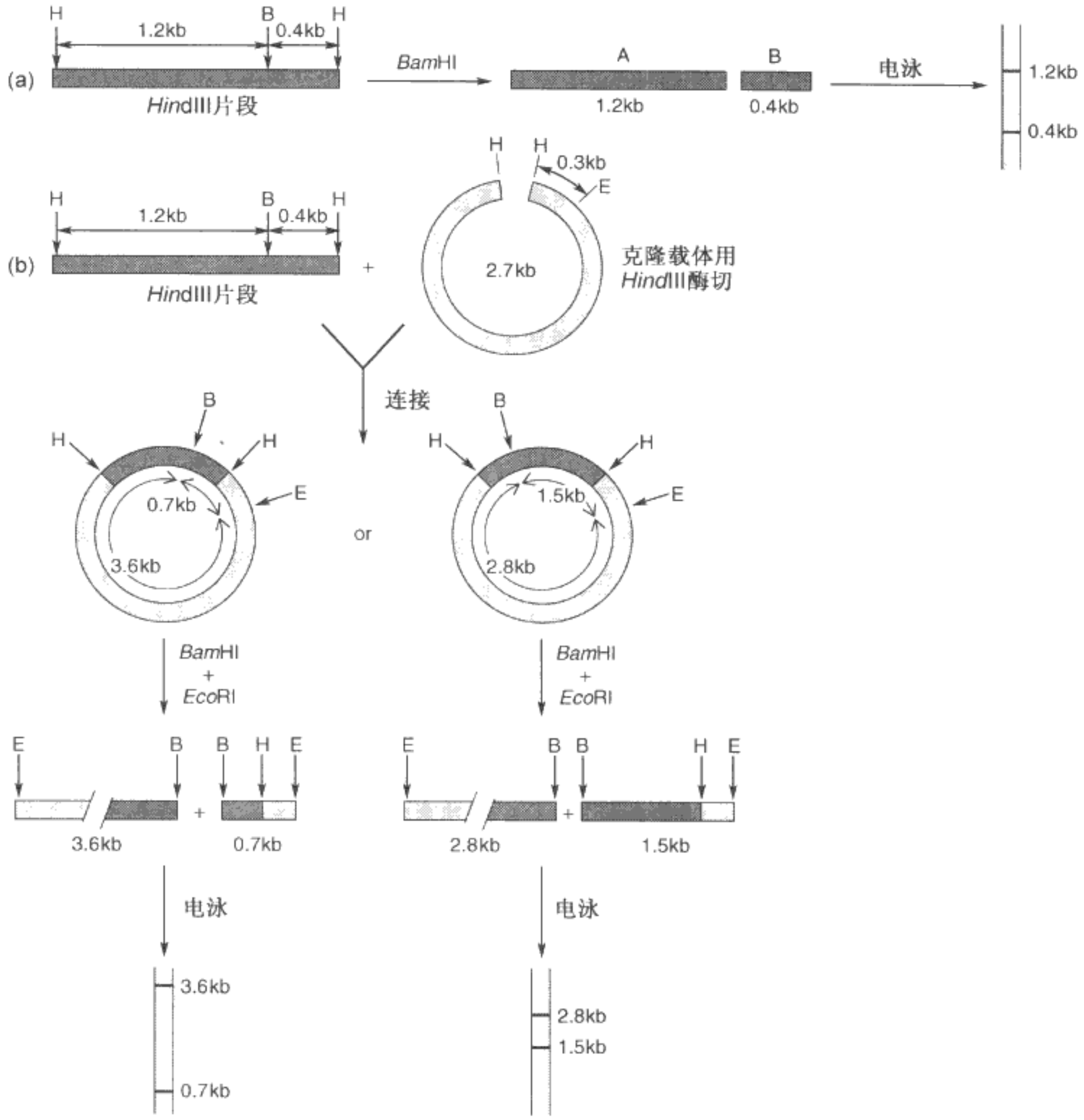


图 5.21　简单的限制酶酶切图谱实验。（a）确定 *Bam*HI 的位点。用 *Bam*HI 切割 1.6kb 的 *Hin*dIII 片段得两个片段，其大小经电泳检测分别为 1.2kb 和 0.4kb 证明 *Bam*HI 切割一次，*Hin*dIII 片段的一端是 1.2kb，另一端是 0.4kb。（b）确定克隆载体上 *Hin*dIII 片段的方向。1.6kb 的 *Hin*dIII 片段可按两种可能的方向插入克隆载体：①*Bam*HI 位点靠近载体的 *Eco*RI 位点；②*Bam*HI位点远离载体的 *Eco*RI 位点。为确定具体方向，同时用 *Eco*RI 和 *Bam*HI 酶切 DNA 电泳检测产物大小。小片段（0.7kb）表明两个位点接近（左图），长片段（1.5kb）说明两个位点较远（右图）。

为回答这个问题，先在载体中确定一个相对于 *Hin*dIII克隆位点为非对称的限制酶位点。在这个例子中，*Eco*RI 位点距 *Hin*dIII 位点仅 0.3kb。这表明如果用 *Bam*HI 和 *Eco*RI 酶切左图中的 DNA 克隆，将会得到大小分别为 3.6kb 和 0.7kb 的两个片段。如果同样用这两种酶切割右图中的克隆，那么会得到 2.8kb 和 1.5kb 两个片段。通过电泳测定酶切片段的大小就能很容易地区别这两种可能性，如图 5.21 底部所示。通常是从几个不同的克隆提取 DNA，每个克隆的 DNA 均用这两种酶切割，然后将这些酶切片段在并排的泳道中电泳，同时留出一个泳道给已知大小的标准片段。平均有一半的重组克隆是一个方向，另一半是另外一个方向。你可以用同样的方法回答下面有关图谱的问题。

问题解答

问题 2

绘制 17kb 的双链环状 DNA 图谱。如图 5.22a

所示，*Hin*dIII 酶切两次，得到两个长度分别为 11kb 和 6kb 的片段（*Hin*dIII-A 和 *Hin*dIII-B）。*Pst*I 也酶切两次，产生两个长度分别为 5kb 和 12kb 的片段（*Pst*I-A 和 *Pst*I-B）。图 5.22b 是以上 4 个片段分别用另一种酶切的结果，如 *Pst*I 酶切 *Hin*dIII-A 片段可产生两个大小分别是 7kb 和 4kb 的片段。

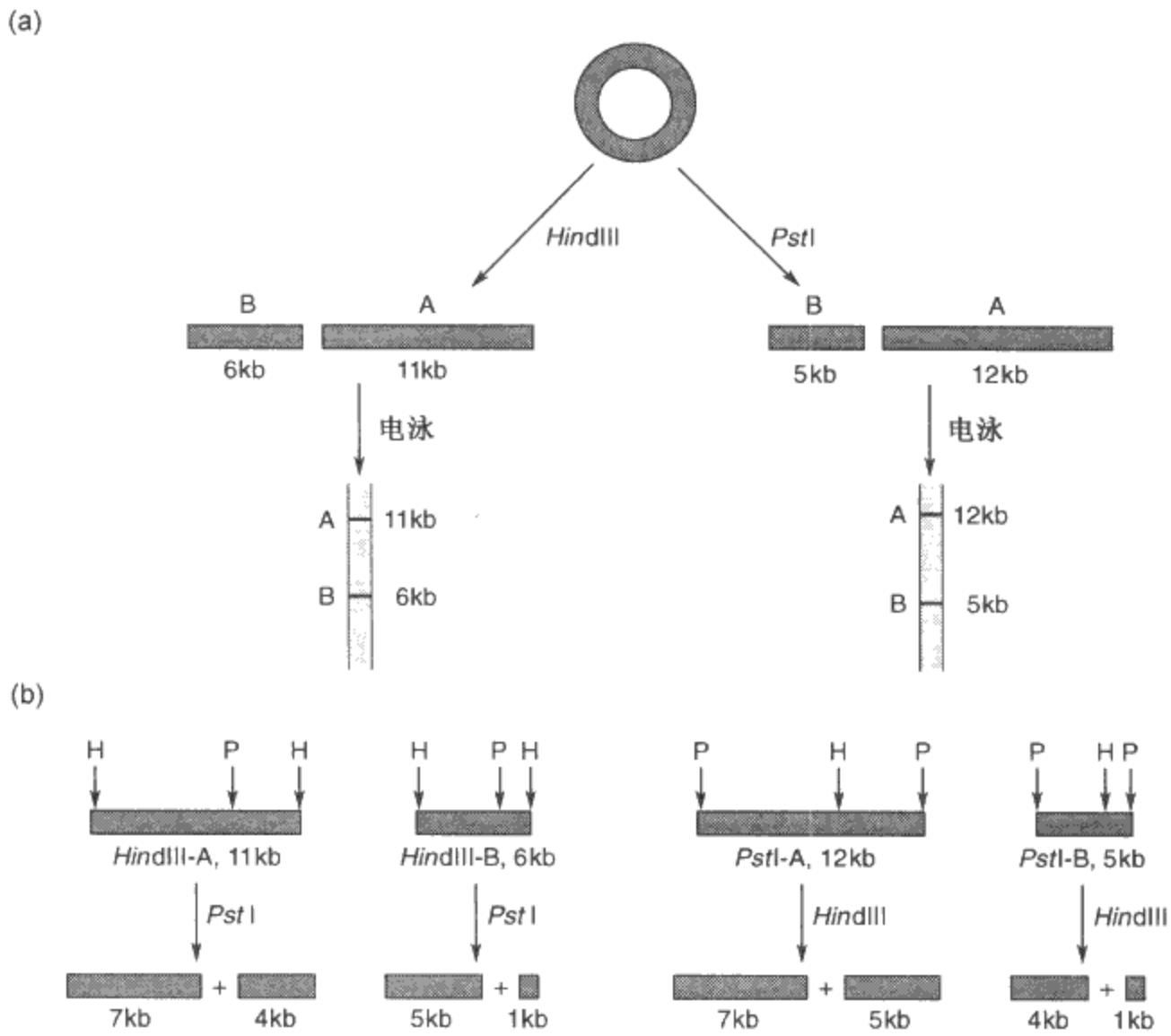


图 5.22　未知 DNA 的限制性内切核酸酶图谱。缩略词：H ＝*Hin*dIII 位点；P＝*Pst*I 位点。

请标出所有 4 个限制酶位点在环状 DNA 上的位置。为简化起见，先将 *Hin*dIII-B 片段放在环的右上角，使一个 *Hin*dIII 位点在 DNA 环的正上方（在 12 点处），另一个 *Hin*dIII 位点在 4 点处，其 *Pst*I 位点靠近 4 点这一端。

解答

两个 *Hin*dIII 片段在 DNA 环状上的位置只有两种可能方式，需要确定哪种是正确的。我们已经人为地将两条 *Hin*dIII 片段中较小的那个放在 DNA 环的右上角，定位其 *Pst*I 靠近 4 点这一端。剩下的唯一问题是 *Hin*dIII-A 的方向，它的 *Pst*I 位点是靠近 *Hin*dIII-B 片段的 12 点处还是 4 点处。要回答这一问题，我们需要考虑 *Pst*I 酶切片段的大小。假设让 *Hin*dIII-A 片段的 *Pst*I 位点靠近 12 点处，如图 5.23a 所示，由于该位点距 *Hin*dIII 位点是 4kb，而 *Hin*dIII-B 上的 *Pst*I 位点距 *Hin*dIII 位点 5kb，所以用 *Pst*I 切割质粒应得到 9kb 的片段。而且，因为质粒全长为 17kb，那么另一条 *Pst*I 片段应为 17－9＝8kb。但是这与观察到的结果不符，因此第二种可能

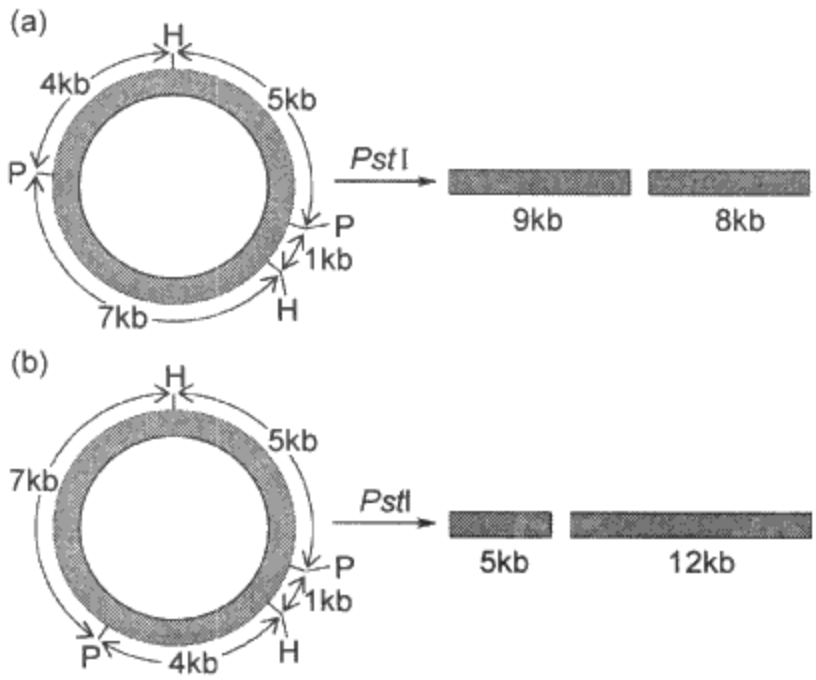


图 5.23　未知 DNA 可能的两个图谱。缩略词：H＝*Hin*dIII 位点；P＝*Pst*I 位点

性一定是对的。来证明一下，如果 *Hin*dIII-A 片段的 *Pst*I 位点靠近 4 点的 *Hin*dIII 位点，如图 5.23b 所示，那么用 *Pst*I 酶切质粒应产生 12kb（7＋5）和

5kb（4＋1）两个片段。这正是实验观察到的，所以这是正确的排列。

这个例子相对简单，但是可用相同的逻辑解决更复杂的图谱问题。有时这种方法有助于标记一段限制性片段（放射性或非放射性标记），并与另一个限制酶酶切片段的 Southern 印迹杂交从而鉴定片段间的相互关系。以图 5.24 所示的线状 DNA 为例，在不使用杂交技术的情况下我们也能推测出酶切位点的顺序，但是这并不容易。考虑仅通过几个杂交就可以获得的信息，如将 *Eco*RI 酶切片段做 Southern 印迹，然后与标记的 *Bam*HI-A 探针杂交，则 *Eco*RI-A 和 *Eco*RI-C 都会被标记，表明 *Bam*HI-A 片段与这两个 *Eco*RI 片段有重叠区。如果将该印迹膜与 *Bam*HI-B 杂交，则 *Eco*RI-A 和 *Eco*RI-D 会被标记，表明 *Bam*HI-B 与 *Eco*RI-A 和 *Eco*RI-D 有重叠。最终通过杂交发现，除了 *Bam*HI-A 和 *Bam*HI-B 以外，没有别的 *Bam*HI 片段与 *Eco*RI-A 杂交，所以 *Bam*HI-A 和 *Bam*HI-B 一定是相邻的。用这种方法，我们可以拼出整个 30kb 片段的物理图谱。

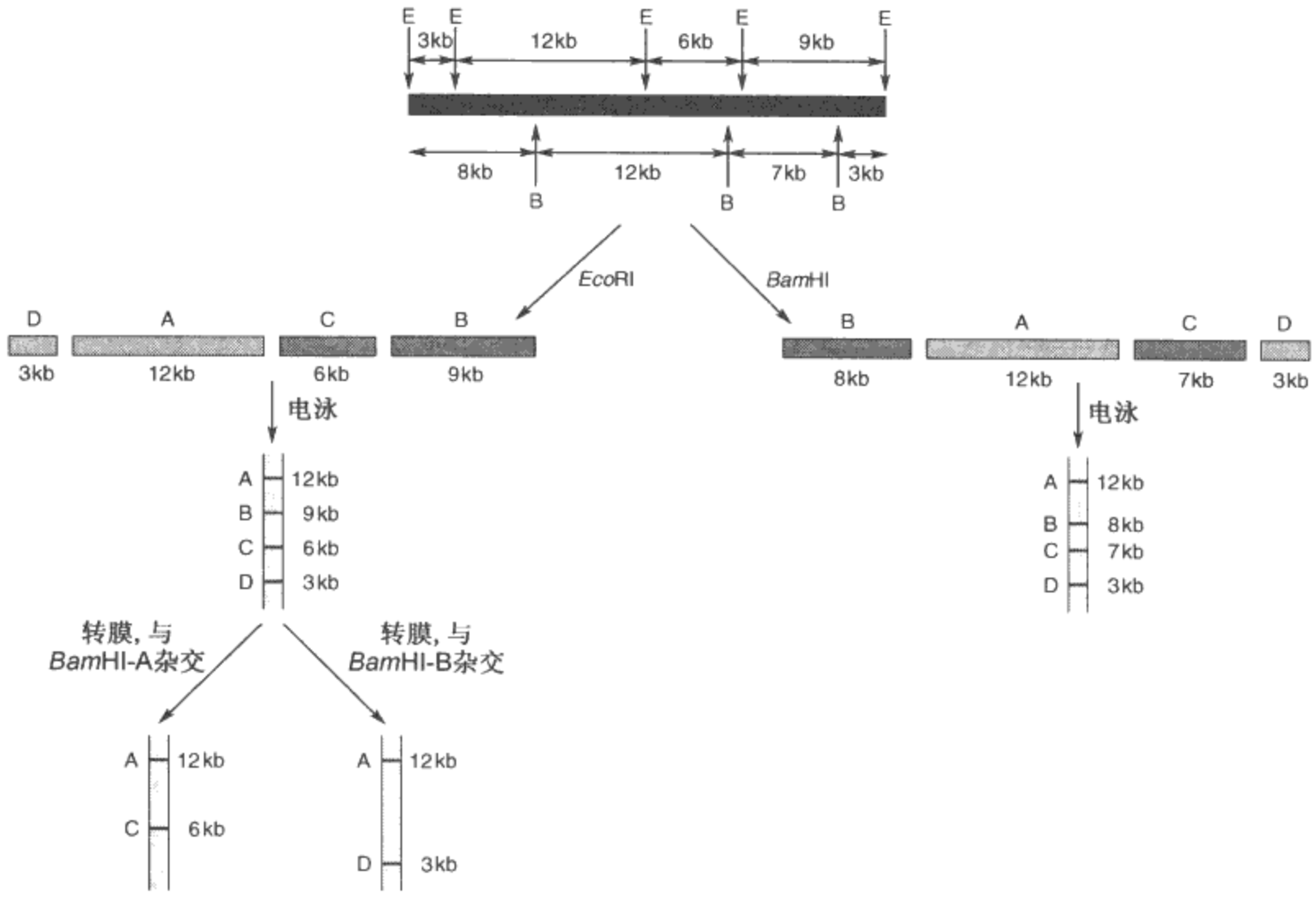


图 5.24　用 Southern 印迹绘制限制性图谱。绘制一个 30kb 片段的图谱，它被 *Eco*RI（E）和 *Bam*HI（B）分别酶切 3 次。为有助于绘图，先用 *Eco*RI 酶切，电泳 4 个酶切片段（*Eco*RI-A、*Eco*RI-B、*Eco*RI-C、*Eco*RI-D），然后 Southern 印迹并与标记的克隆的 *Bam*HI-A 和 *Bam*HI-B 片段杂交。左下方显示的结果说明 *Bam*HI-A 片段与 *Eco*RI-A 和-C 重叠，*Bam*HI-B 片段与 *Eco*RI-A 和 *Eco*RI-D 重叠。这些信息辅以 *Eco*RI 片段的 *Bam*HI 酶切结果（反之亦然）可拼接出完整的酶切图谱。

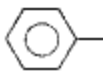
小结　物理图谱是关于 DNA 分子上的物理“坐标”，如限制性酶切位点等的空间排列顺序。绘制限制性图谱（对限制性位点的作图）的一个主要策略是用两种或两种以上的限制酶分别酶切所研究的 DNA，然后测定每个酶切片段的大小，再用另一种限制酶切割这些酶切片段，用凝胶电泳分别测定第二次酶切片段的大小。由此至少可确定一些识别位点与其他识别位点的相对位置。还可以有效地改进这一过程，做法是对某些酶切片段做 Southern 印迹，然后与被标记的另一种酶的酶切片段杂交，这一方法可以显示两个酶切片段之间是否有重叠。

基于克隆基因的蛋白质工程：定点诱变

通常，生化学家习惯用化学的方法改变所研究蛋白质的某些氨基酸，以便观察这些改变对蛋白质活性的影响。但是，化学试剂作为蛋白质操作的工

具太困难了，很难保证只有一个或一类氨基酸发生改变。利用克隆的基因能使这类研究更精确，允许我们对蛋白质进行显微操作。通过改变基因中的特定碱基，就能改变该基因蛋白质产物中相应位点的氨基酸，然后观察这些改变对蛋白质功能的影响。

假设已经克隆了一个基因，想改变该基因的一个密码子。具体来讲，这个基因编码的氨基酸序列包含一个酪氨酸，而酪氨酸含有一个酚基：



为了研究这个酚基的重要性，将酪氨酸的密码子变成苯丙氨酸的密码子。苯丙氨酸与酪氨酸类似，但是它所含的基团不是酚基而是苯基：

如果酪氨酸的酚基对蛋白质的活性很重要，那么用苯丙氨酸的苯环代替酚基时，蛋白质的活性就会减弱。

在该例中假设要将密码子 TAC（Tyr）替换成 TTC（Phe），怎样做这个**定点诱变**（site-directed mutagenesis）呢？图 5.25 介绍了一种常用的基于 PCR 的定点诱变技术。先从克隆一个含有酪氨酸密码子（TAC）的基因开始。DNA 上的 CH_3 符号表示该 DNA 分子与来自大多数 *E. coli* 菌株的 DNA 一样，在 5′-GATC-3′序列上发生了甲基化。这段甲基化的序列恰好是限制酶 *Dpn*I 的识别位点，这一位点的作用将在后面讲述。图上只显示了两个甲基化的 *Dpn*I 位点，而实际上有很多，在一段随机 DNA 序列中大约每 250bp 就会出现一个 GATC。

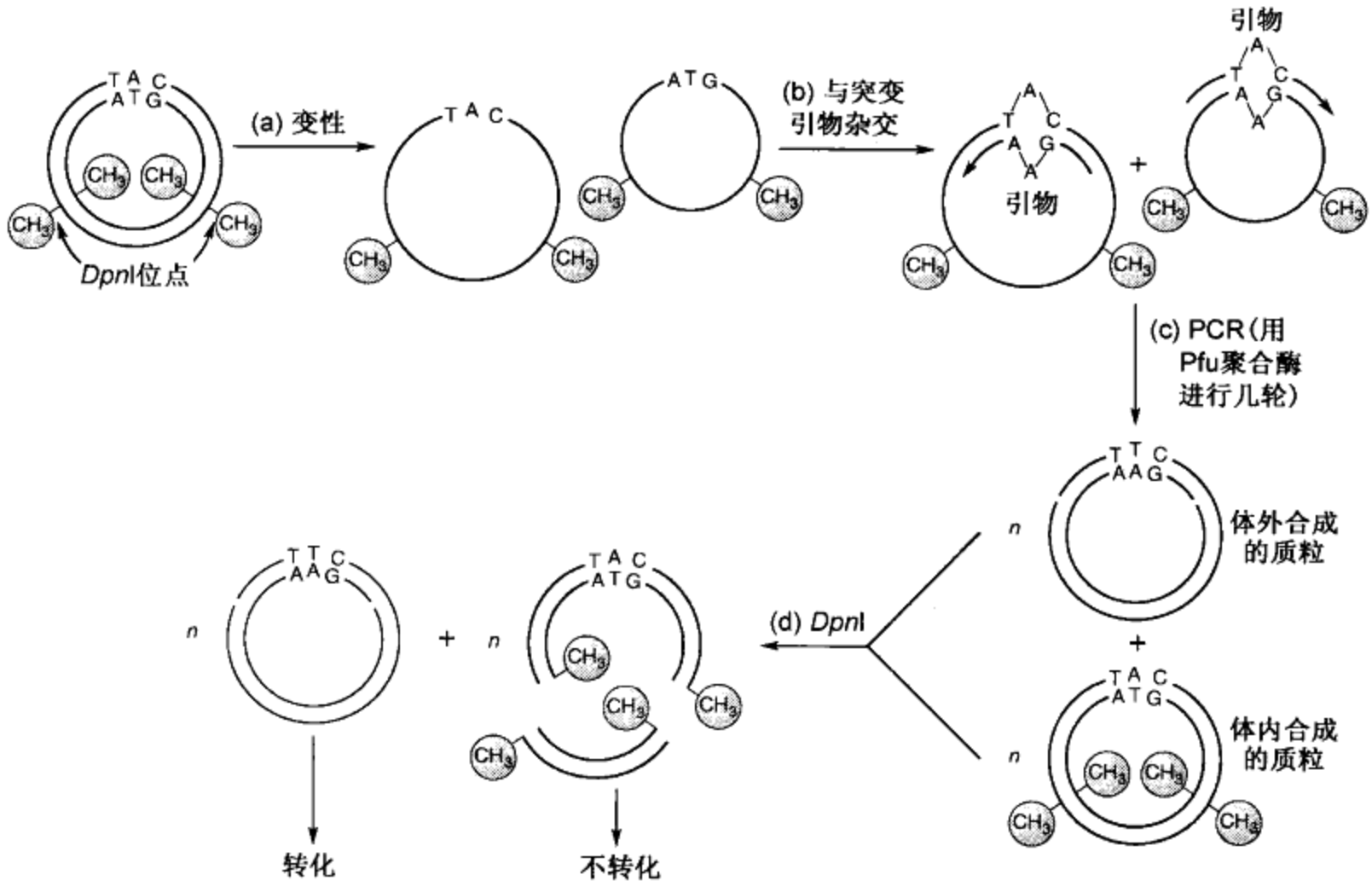


图 5.25　基于 PCR 的定点诱变。从含有酪氨酸密码子 TAC 基因的质粒开始，其中 TAC 将被突变为苯丙氨酸密码子 TTC。因此，原来的 A-T 碱基对要变为 T-A 碱基对。该质粒是从正常的 *E. coli* 菌株分离的，它能对 GATC 序列的 A 甲基化（—CH_3 表示）。（a）加热使质粒双链分离。本来质粒的两条链相互缠绕不能完全分开，为简单起见在这里显示为完全分离。（b）用含 TTC 或其反向互补密码子 GAA 的突变引物与单链 DNA 杂交，每条引物的突变碱基用红色表示。（c）用突变引物进行几轮（约 8 个）PCR 反应，扩增密码子突变的质粒。用高保真的热稳定性 DNA 聚合酶（如 Pfu 聚合酶），可尽量降低质粒扩增中出现的错误。（d）用 *Dpn*I 处理 PCR 反应中的 DNA 以消化甲基化的野生 DNA，而 PCR 产物是体外合成的没有甲基化，所以不会被降解。最后用处理过的 DNA 转化 *E. coli*，理论上只有突变的 DNA 能够转化。通过对几个克隆的质粒进行测序可检测转化的情况。

第一步是将 DNA 热变性，第二步是将突变引物与 DNA 杂交。其中一个 25bp 的寡核苷酸引物（25mer）序列如下：3′-CGAGTCTGCCAAAGCATGTATAGTA-5′。

引物设计成与基因的一段非模板链序列一致，只是将中间的 ATG 三联体密码子换成了 AAG，改变的碱基用下划线标出，另一个引物是互补的 25mer。两个引物都含有替换的碱基以便改变目标密码子。第三步是用这对引物进行几轮 PCR 扩增 DNA 片段，掺入了想要改变的碱基。我们只进行几轮 PCR 循环以减少在 DNA 复制中出现的其他偶然性突变。为同一目的，使用高保真的 Pfu DNA 聚合酶。这种酶是从古菌 *Pyrococcus furiosus*（拉丁名：*furious fireball*）中纯化得到的，该菌生活在海底火山口附近的沸水中。Pfu DNA 聚合酶对其所合成的 DNA 具有校正功能，因此很少出错。另一种与之相似的酶叫做 vent 聚合酶（vent polymerase），来自另一种超耐热极嗜热古菌。

得到突变 DNA 后，需要将其与剩余的野生型 DNA 分开或者将后者降解。前面提到的野生型 DNA 甲基化在这里就派上用场了。*Dpn*I 酶只对甲基化的 GATC 位点进行切割。由于野生型 DNA 有甲基化而体外合成的突变型 DNA 没有甲基化，所以 *Dpn*I 酶只切野生型 DNA。一旦被酶切，野生型 DNA 就不能再转化到 *E. coli* 中，只有突变型 DNA 可以转化。也可对多个克隆进行测序，以确保是突变型而不是原来的野生型序列，一般情况下都是突变型的。

小结 利用克隆的基因，可以按照意愿引入突变，从而改变蛋白质产物的氨基酸序列。可以利用双链 DNA、两条互补的突变引物和 PCR 得到突变 DNA。只需用 *Dpn*I 消化 PCR 产物就可除去几乎所有的野生型 DNA，这样就可以用突变的 DNA 转化宿主细胞了。

5.4 转录物的图谱定位与定量

分子生物学经常讨论的一个主题是转录图谱的定位（定位转录的起始和终止位点）及其定量（测定在某一时刻转录了多少）。分子生物学家利用多种技术进行转录图谱定位和定量，我们将在本书介绍几种。

测定某一时刻产生了多少转录物的最简单方法是标记转录物，在体内或体外将标记的核苷酸掺入转录物中，然后电泳，通过放射性自显影检测凝胶上的转录条带。实际上已经用这种方法在体内和体外都做过转录物的分析。但是在体内只有当所研究的转录物相当丰富且电泳时容易与其他 RNA 分开时，这种方法才有效果。tRNA 和 5S rRNA 都满足以上两个条件，它们在体内的合成情况已经用简单的电泳进行了示踪检测（第 10 章）。而在体外，这种直接方法只适合具有明确转录终止子的转录物，这样才能呈现分开的 RNA 条带，否则就会因具有不同 3′端的 RNA 片段而出现无法区分的弥散涂抹带。在原核生物转录中这种情况时有发生，但在真核的例子很少见。因此常需借助其他的方法，这些方法虽不直接但更为专一。

有几种常用技术用于转录物的 5′端图谱定位，其中一种还能对 3′端进行定位，有些方法还能定量分析细胞在特定时刻基因的转录物。这些方法依赖于核酸的杂交能力，以便从数千个 RNA 中检测出某种 RNA。

Northern 印迹

假设已经克隆到一个 cDNA（RNA 的 DNA 拷贝），想知道相应的基因（基因 X）在生物体 Y 的各个不同组织中的表达活性怎样。有多种方法可回答这个问题，但是这里介绍的方法还能回答该基因产生的 mRNA 的大小。

首先提取所研究生物不同组织的总 RNA，然后用琼脂糖凝胶电泳分离 RNA 并印迹到一个合适的介质上。鉴于与之类似的 DNA 印迹叫做 Southern 印迹，所以很自然地将 RNA 印迹叫做 **Northern 印迹**（Northern blot）。

接下来，将 Northern 印迹膜与标记的 cDNA 探针杂交。印迹膜上与探针互补的 mRNA 出现在哪里，杂交就在哪里发生，所产生的带标记条带可用 X 线胶片检测。如果未知 RNA 旁边的泳道跑上已知大小的标准 RNA，就可以知道与探针杂交“发亮”的 RNA 条带的大小。

同时，Northern 印迹还告诉我们 X 基因转录物

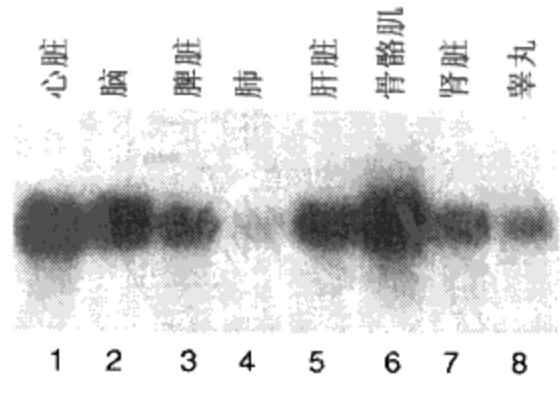


图 5.26 Northern 印迹膜。从大鼠不同组织（在图的上部标出）分离的细胞质 mRNA，然后取各种组织的等量 RNA 电泳，再进行 Northern 印迹。将印迹膜上的 RNA 与标记的大鼠 3-磷酸甘油醛脱氢酶（G3PDH）基因探针杂交，然后杂交膜与 X 线胶片曝光。条带表示 G3PDH 的 mRNA，其亮度指示每种组织中该 mRNA 的表达量。（*Source*：Courtesy Clontech.）

的丰度，条带所含 RNA 越多，与之结合的探针就越多，曝光后胶片上的条带就越黑。通过密度计测量条带的吸光度可定量条带的黑度，或用磷屏成像法直接定量条带上标记的量。图 5.26 是大鼠 8 种不同组织 RNA 的 Northern 印迹，用探针与印迹膜杂交寻找编码糖代谢相关的 G3PDH（3-磷酸苷油醛脱氢酶）的基因。很明显该基因的转录物在心脏和骨骼肌最丰富，而在肺中最少。

小结 Northern 印迹与 Southern 印迹类似，但是它含的是电泳分离的 RNA 而不是 DNA，印迹上的 RNA 可通过与标记的探针杂交进行检测，条带的亮度显示的是其中特异 RNA 的相对量。

S1 图谱定位

S1 图谱定位（S1 mapping）技术用于定位 RNA 的 3′端和 5′端及定量细胞中给定时间的特定 RNA。其原理是标记一个仅能与目的 RNA 杂交的单链 DNA 探针，探针必须跨转录物的起始或终止区。在探针与转录物杂交后，加入只降解单链 DNA 和 RNA 的 **S1 核酸酶**（S1 nuclease）。这样，探针与转录物杂交部分受到保护而免于降解。杂交部分的大小可通过电泳检测出来，并且受保护的范围能告诉我们转录起始和终止的位置。图 5.27 详细显示了如何用 S1 图谱定位转录起始位点。首先，用 ^{32}P 标记 DNA 探针的 5′端。通常，DNA 的 5′端含有非放射性磷酸基，在添加放射性磷酸之前需要用碱性磷酸酶将其去除，然后用多核苷酸激酶将［γ-^{32}P］ATP 的磷酸基团转移到 DNA 链起始端的 5′羟基上。

在本例中，*Bam*HI 片段的两端均被标记，可产生两个标记的单链探针。但是这样会干扰分析，因此左端的标记必须去除，用另一种限制酶 *Sal*I 再切一次 DNA，然后凝胶电泳将短的左侧片段与用作探针的长片段分开。现在双链 DNA 仅在一端被标记，可在变性后产生一个带标记的单链探针。下一步，DNA 探针与含有目的 RNA 的细胞 RNA 混合液杂交，探针与互补转录物的杂交会在左侧留下一个单

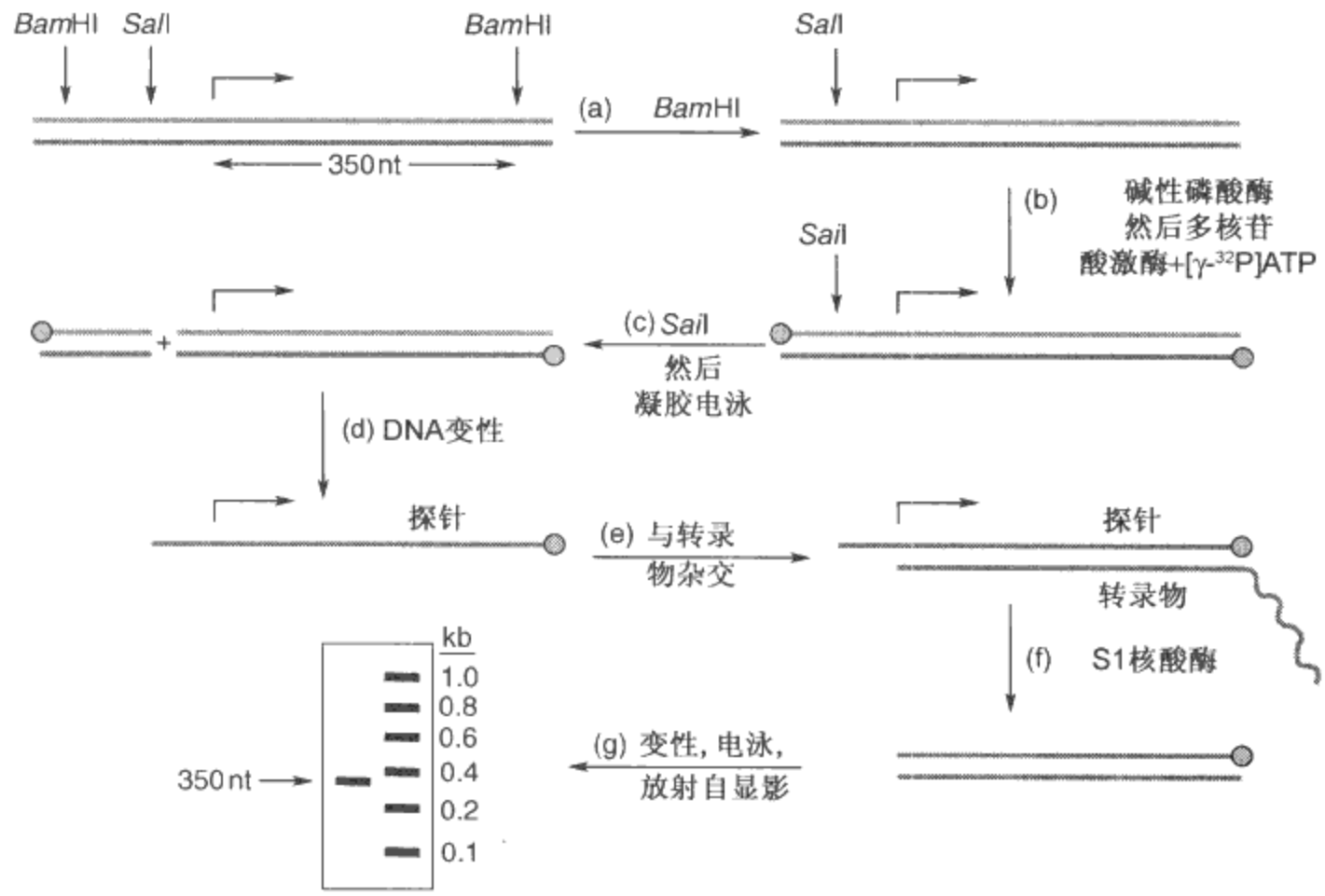


图 5.27 S1 图谱法定位转录物的 5′端。从克隆的带有几个已知限制酶位点的双链 DNA 开始。现在确切的转录起始位点还不知道，通过 S1 图谱法要确定的位置先标记为（⌜）。已知转录起点两侧有 *Bam*HI 位点，一个 *Sal*I 位点在起始位点上游。步骤（a），*Bam*HI 酶切产生 *Bam*HI 片段如图右上方所示。步骤（b），除去该片段 5′羟基上未标记的磷酸基，然后用多核苷酸激酶和［γ-^{32}P］ATP 标记 5′端，圆圈表示标记的末端。步骤（c），用 *Sal*I 酶切，电泳分离切出的两个片段，除去左端的标记。步骤（d），将此双链 DNA 变性产生单链探针与（e）中的转录物杂交。步骤（f），用 S1 核酸酶处理杂交产物，消化杂交体中左端的单链 DNA 和右端的单链 RNA。步骤（g），将杂交体变性，电泳检测被保护探针的长度，在另一泳道加入已知大小的 DNA 片段作为标准。受保护区探针的长度指示了转录起始位点的位置。本实验中，转录起始位点位于探针中被标记的 *Bam*HI 位点上游 350bp 处。

链的 DNA 尾，右侧留下单链的 RNA 尾。接下来用 S1 核酸酶处理，特异性地降解单链 DNA 或 RNA，使双链多核苷酸包括 RNA-DNA 杂交体保留完整。这样，与转录物杂交的 DNA 探针部分包括末端标记被保护起来。最后，探针受保护 这部分的长度可通过高分辨率的凝胶电泳并与标准 DNA 比较而确定。由于探针右端的定位已经确切知道（标记的 *Bam*HI 位点），受保护这部分探针的长度会自动告诉我们探针左端，即转录起始位点的定位。在该例中受保护探针 350nt 长，因此转录起始位点位于标记的 *Bam*HI 位点上游 350bp 处。

也可用 S1 图谱定位技术定位转录物的 3′端。如图 5.28 所示，将 3′端标记的探针与目的 RNA 杂交，其他步骤均与前面所讲的 5′端图谱定位相同。但是 3′端标记与 5′端标记不同，因为 5′端标记所用的多核苷酸激酶不能使核酸分子的 3′羟基磷酸化。标记 3′端的一种方法是**末端补平法**（end-filling），如图 5.29 所示。用产生 3′端缺损的限制酶切割 DNA 分子，缺损的 3′端可在体外延伸到与 5′端齐平。如果在末端补平反应中使用有标记的核苷酸，那么 DNA 的 3′端就可标记上。

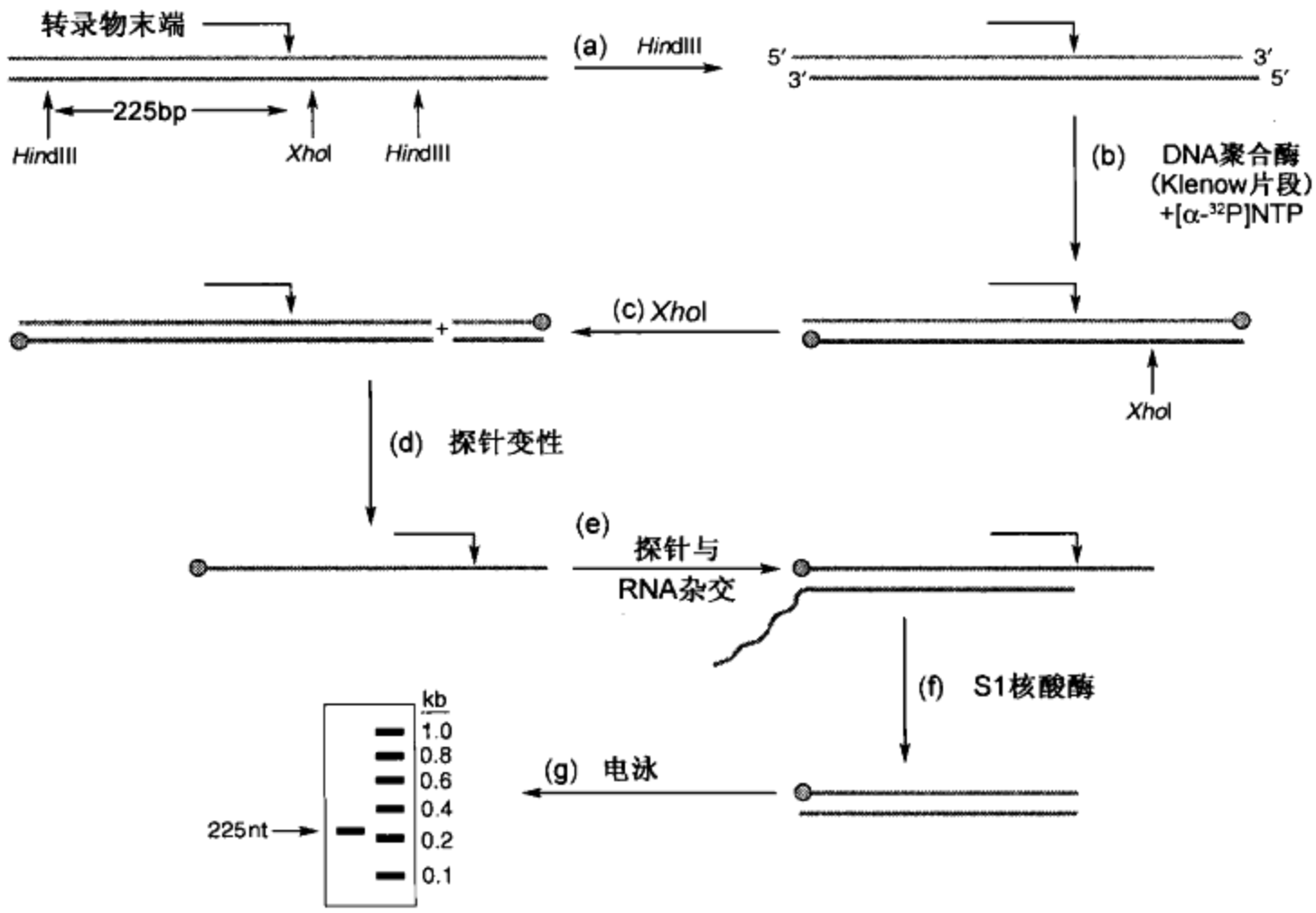
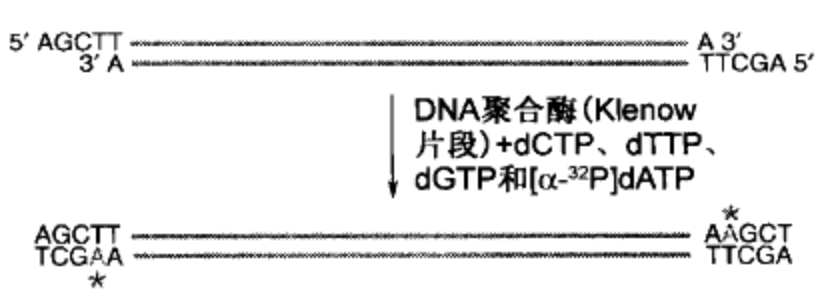


图 5.28　S1 图谱法定位转录物的 3′端。原理与 5′端图谱定位类似，不同的是在 3′端而不是 5′端标记探针（详见图 5.29）。步骤（a），用 *Hin*dIII 酶切，然后在步骤（b）标记酶切片段的 3′端，圆圈表示标记的末端。在步骤（c）用 *Xho*I 酶切，凝胶纯化左端标记的片段。步骤（d）探针变性，在步骤（e）与 RNA 杂交。步骤（f）用 S1 酶去除探针（及单链 RNA）的未保护区。最后，步骤（g）电泳带标记的被保护探针，检测其大小，本实验是 225nt，表明转录物的 3′端位于探针左侧标记的 *Hin*dIII 位点下游 225bp 处。

图 5.29　末端补齐法标记 DNA 3′端。顶部的 DNA 片段用 *Hin*dIII 酶切产生，如图所示两端均有 5′突出端，可用 Klenow 片段（见第 20 章）DNA 聚合酶填充补齐。这个酶优于完整的 DNA 聚合酶之处在于它缺乏正常的 5′→3′外切核酸酶活性，而这一活性会使 5′黏性末端在填补之前被降解。末端补齐反应以 4 种核苷酸为底物，其中 dATP 带有标记，以便标记 DNA 末端。若要标记更强，可以用一种以上带标记的核苷酸。



S1 图谱定位不仅用于转录物末端的定位，还能确定转录物的浓度。假设探针是过量的，那么放射自显影条带的强度就与保护探针的转录物的浓度呈正比，转录物越多，受保护的探针就越多，放射自显影条带就越强。因此，只要知道与目的转录物相对应的条带，其强度就可用于测定转录物的浓度。

从 S1 图谱定位理论衍生出来的一种重要的技术叫 **RNase 作图**（RNase mapping）也叫 **RNase 保护分析**（RNase protection assay）。这个方法与 S1 图谱定位类似，同样能提供特定转录物 5′端和 3′端定位及其浓度方面的信息。但是该方法使用 RNA 探针，因此被 RNase 而不是 S1 核酸酶降解。这种技术很流行，因为 RNA 探针（riboprobe）的制备相对容易，可在体外用纯化的噬菌体 RNA 聚合酶对重组质粒或噬菌粒进行转录获得。RNA 探针的另一个优势是通过向体外转录反应中加入一种带标记的核苷酸，就能被标记成极高的活性，产生均一标记的而不是末端标记的探针。探针的专一活性越高，检测微量转录物的灵敏度就越高。

小结 S1 图谱定位技术利用标记的 DNA 探针探测转录物的 5′端或 3′端。探针与转录物的杂交保护了部分探针免受专门降解单链多核苷酸的 S1 核酸酶的降解。转录物所保护部分探针的长度可定位转录物的末端。由于被转录物保护探针的量与转录物浓度呈正比，S1 图谱定位也可用作定量的方法。RNase 图谱定位是 S1 图谱定位的衍生方法，它用的是 RNA 探针和 RNase，而不是 DNA 探针和 S1 核酸酶。

引物延伸

S1 图谱定位已经应用在一些经典实验中，我们将在后面几章中介绍，它是转录物 3′端定位的最好方法，但也存在不足之处。一是 S1 核酸酶在 RNA-DNA 杂合体的末端，甚至在 RNA-DNA 杂合体的富 A-T 区瞬时解链时会“咬掉”一点点。另一方面，有时候 S1 核酸酶不能完全降解单链部分，使转录物看起来比其实际长度稍长。假如对核酸末端的定位要精确到单个核苷酸，那么上述这些缺点就很严重了。另外一种叫做**引物延伸**（primer extension）的方法可以把 5′端（不是 3′端）精确定位到单个核苷酸水平。

图 5.30 给出了引物延伸法的工作原理。第一步转录，通常在体内自然发生。只需简单地收集包含目的转录物的细胞 RNA 粗提液，转录物的序列已知，5′端有待定位。接着，将标记的约 18nt 的寡核苷酸（引物）与细胞 RNA 杂交。注意，就像 S1 图谱定位的特异性来自探针与目的转录物的互补性一样，该方法的特异性也来自引物与目的转录物的互补性。原则上，这个引物（或 S1 探针）能够从众多不相关的 RNA 中挑出我们想要定位的转录物。

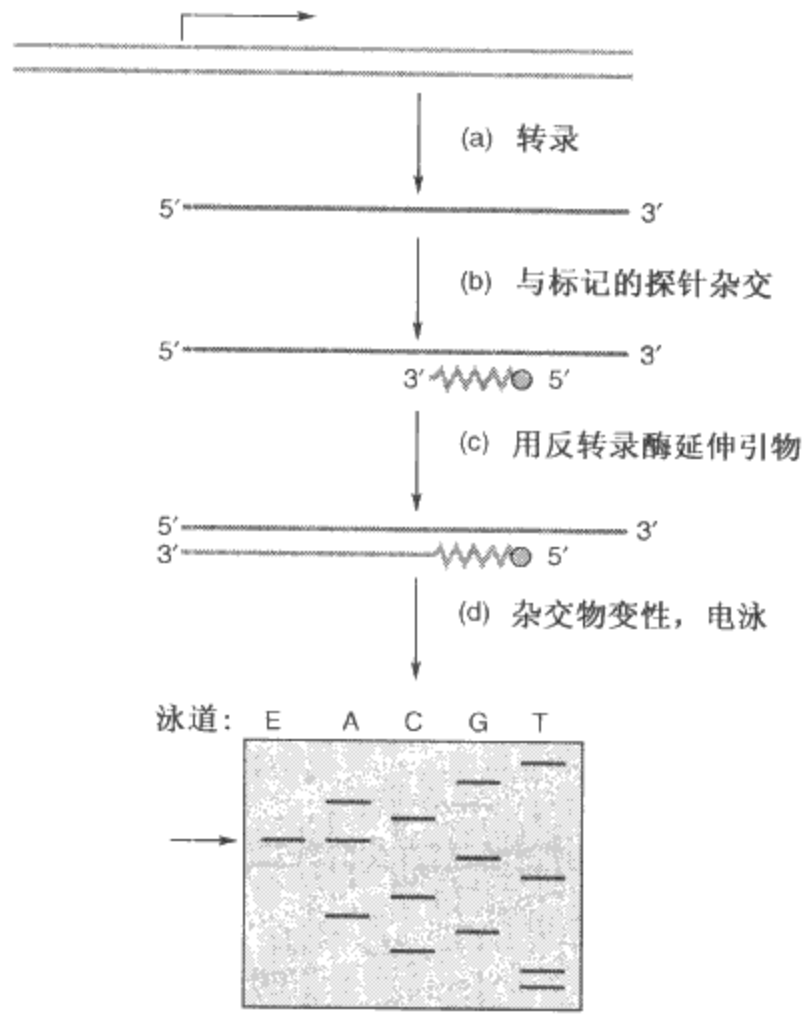


图 5.30 引物延伸法。（a）细胞内自然发生转录，因此从获得细胞 RNA 开始。（b）已知转录物的部分序列，合成并标记一段 DNA 寡核苷酸引物，它与推测的 5′端不太远的区域互补。将这段引物与转录物杂交，它应该特异性地与该转录物而非其他序列杂交。（c）用反转录酶合成与转录物互补的 DNA 链，将引物一直延伸到 RNA 的 5′端。如果引物自身未被标记或希望对延伸引物引入额外的标记，则可在这一步加入标记的核苷酸。（d）将杂交体变性并电泳标记的延伸引物（泳道 E）。在其他泳道（泳道 A、C、G、T）以相同的引物及转录区 DNA 进行测序反应作为标准。理论上这样可以指出转录起始位点的精确碱基，本例中延伸引物（箭头所指）与 A 泳道的一个 DNA 片段共迁移。因为在引物延伸反应和所有的测序反应中使用的引物相同，说明转录物的 5′端与 TTCGACTGACAGT 序列中间的 A（加下画线的）相对应。

下一步，用反转录酶将寡核苷酸引物向目的转录物的 5′端延伸。如第 4 章所述，反转录酶是一种逆向进行转录反应的酶，即以 RNA 为模板复制 DNA。因此，该酶正好适合我们要求它做的工作：合成待定位 RNA 的 DNA 拷贝。引物延伸反应完成后，将 RNA-DNA 杂合体变性并将标记的 DNA 与

标准 DNA 分子在高分辨率（如 DNA 测序凝胶）上电泳。实际上很方便用引物延伸反应的引物和克隆的 DNA 模板一起进行一系列测序反应，然后以测序反应的产物作为标准。在本例中，测序反应产物与 A 泳道的一个条带共迁移，表明转录物的 5′端碱基对应序列 TTCGACTGACAGT 中的第 2 个 A（加下画线的），这是对转录起始位点十分精确的定位。

如同 S1 图谱定位技术，引物延伸法也能估算某种转录物的含量。转录物含量越高，与之杂交的标记引物就越多，因此合成的带标记的反转录物也就越多。标记的反转录物越多，电泳凝胶的放射性自显影条带就越黑。

小结 用引物延伸法可以定位转录物的 5′端，做法是将寡核苷酸引物与目的 RNA 杂交，在反转录酶作用下向转录物的 5′端延伸引物，最后电泳测定反转录物的大小。通过这种方法获得的信号强度可测定转录物的含量。

截断转录和无 G 盒转录

如果想要突变一个基因的启动子，观察突变对转录的精确性和效率的影响，必须有一种方便的分析方法以提供两方面的信息：①转录是否精确（如引物延伸法或其他实验已经定位的转录起始位点）；②这种精确转录的量有多少。可以采用 S1 图谱定位或引物延伸法，但是这两种方法相对复杂。一种更简单的叫**截断转录**（run-off transcription）的方法能很快给出答案。

图 5.31 给出了截断转录的原理。先从一个含有待转录基因的 DNA 片段开始，然后用一种限制性内切核酸酶在转录区的中间切割。接着，用标记的核苷酸在体外转录这个切断的基因，这样转录产物就会被标记。由于切割是在基因的中间，因此聚合酶到达此片段末端后就会“跑脱”（run off），所以这种方法叫做截断转录。现在可以测定截断转录物的长度了。因为事先准确地知道限制酶酶切位点的位置（本例中是 *Sma*I 位点），因此截断转录物的长度（本例中是 327bp）就证实了转录起始位点是在 *Sma*I 位点上游 327bp 处。

注意 S1 图谱定位和引物延伸法很适合体内定位转录物，相比之下截断转录依赖于体外转录，因此它只适合用于在体外能被精确转录的基因，因而不能给出细胞转录物浓度的信息，是测定体外转录效率的好方法。产生的转录物越多，截断转录物的信号就越强。实际上截断转录作为定量方法是最有用的。用 S1 图谱定位或引物延伸法鉴定出生理的转录起始位点后，就可以在体外应用截断转录了。

一种从截断转录衍生出来的精确定量体外转录的技术叫**无 G 盒转录**（G-less cassette transcription）（图 5.32）。该方法不对基因实施酶切，而是将无 G 盒，或非模板链上一段缺少 G 的核苷酸直接插入启动子的下游。模板在体外转录时用 CTP、ATP、UTP，其中一种碱基是标记的，但无 GTP。因此转录会在无 G 盒末端第一个需要 G 的地方停下来，产生大小可预测的中断转录物（基于无 G 盒的大小，通常是几百个碱基对）。电泳转录物，放射自显影检测基因的转录活性。启动子越强，产生的中断转录物越多，放射性自显影相应的条带就越强。

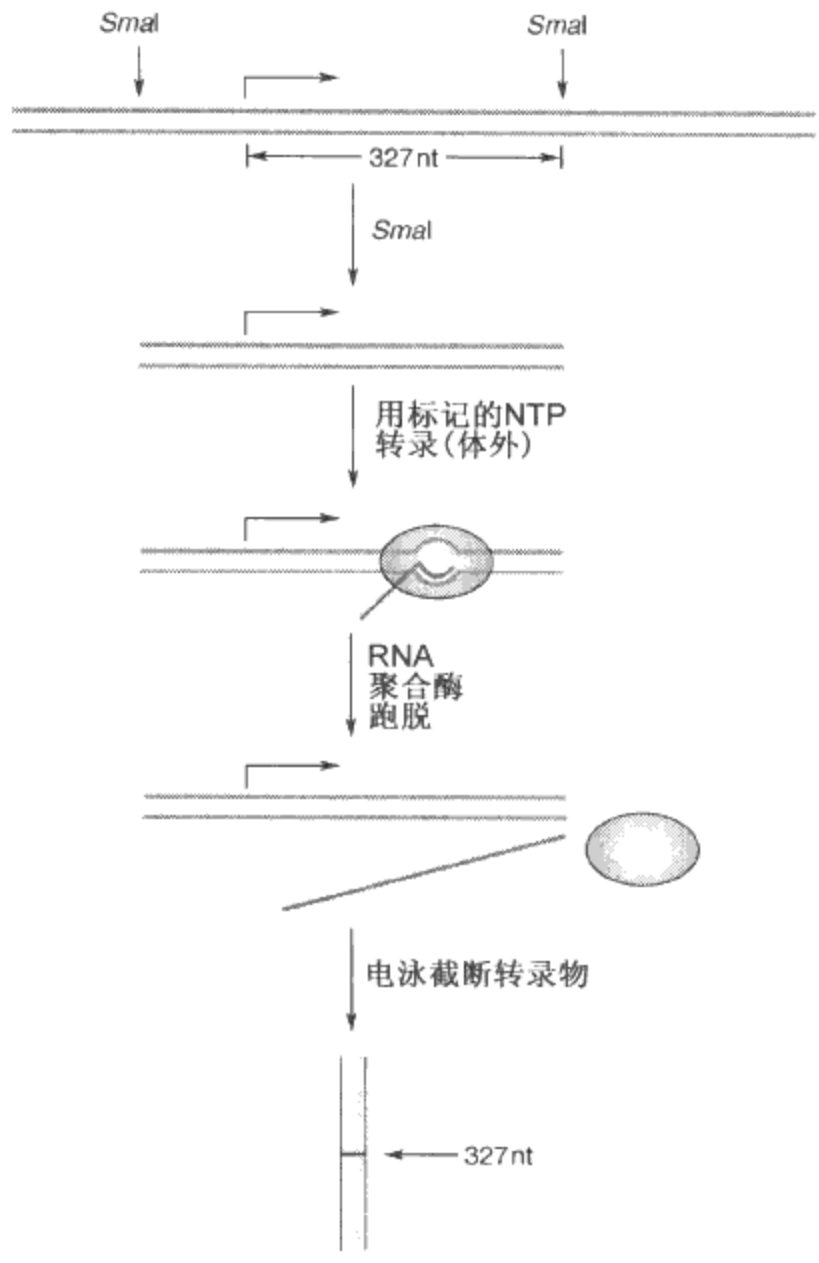


图 5.31 截断转录。从酶切克隆的基因开始，该基因的转录待测定。对该截短的基因进行体外转录。当 RNA 聚合酶（椭圆形）到达截短基因的末端时脱落下来，释放截断转录物，其大小（本例是 327nt）可通过凝胶电泳测定，与转录起始位点和截短基因 3′端已知酶切位点（本例是 *Sma*I 位点）间的距离相对应。基因的转录活性越高，产生的 327nt 片段的信号越强。

小结 截断转录是一种精确和有效的检测体外转录方法。基因在中间被切开后，在体外用标记的寡核苷酸进行转录，RNA 聚合酶在末端跑

脱，释放出不完整的转录物。截断转录物的大小可定位转录起始位点，转录物的量反映了转录的效率。在无 G 盒转录中，将启动子与非模板链缺失一串 G 的双链 DNA 融合，然后在无 GTP 的条件下进行体外转录。转录在无 G 盒的末端停止，产生在电泳凝胶上大小可预测的条带。

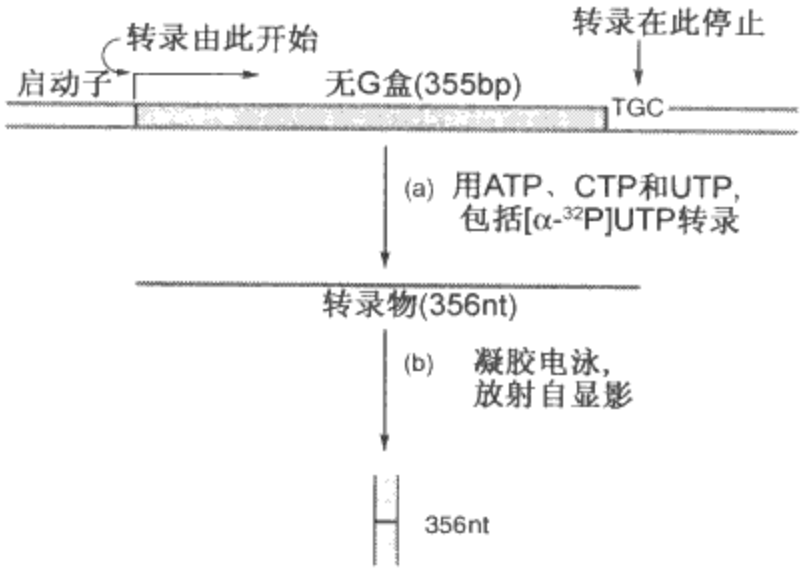


图 5.32 无 G 盒转录。（a）在体外无 GTP 条件下转录无 G 盒（粉色）模板，无 G 盒区插入启动子的下游。该区含 355bp，非模板链不含 G，紧接其后的是 TGC，所以转录在 G 之前停止，产生 356nt 的转录物。（b）电泳标记的转录物，放射自显影测定信号强度，由此表明该盒的转录强度。

5.5 检测体内转录效率

引物延伸、S1 图谱定位和 Northern 印迹都可用来测定特定时刻细胞中某一特定转录物的浓度，但却不一定能告诉我们转录物的合成效率，因为转录物的浓度不但取决于其合成速率还取决于其降解速率。可以使用其他方法检测转录效率，如核连缀转录分析和报告基因表达等。

核连缀转录分析

该分析（图 5.33a）的思路是从细胞中分离出细胞核，然后在体外继续进行体内已经开始的转录。这种在分离的细胞核中持续进行的转录叫做**连缀转录**（run-on transcription），因为在体内已经启动转录的 RNA 聚合酶在体外仍然“连续”或继续延伸相同的 RNA 链。通常用标记的核苷酸为底物进行连缀反应使产物带上标记。在分离的细胞核中一般不会起始新的 RNA 的合成，因此可以很自信地认为分离的细胞核中发生的任何转录都只是体内已经起始转录的延续，所以连缀反应产生的转录物不但能揭示转录的效率而且还能告诉我们哪些基因在体内发生了转录。为了消除体外新 RNA 链起始合成的可能性，可添加阴离子多糖肝素，它能与游离的 RNA 聚合酶结合阻止转录的再起始。

一旦产生了带标记的连缀转录物就必须进行鉴定，因为转录物很少完整，其大小没有意义。最容易的鉴定方法是**斑点印迹**（dot blotting）（图 5.33b）。将已知样品 DNA 变性后点在膜上，然后将这个**斑点膜**（dot blot）与标记的连缀转录物 RNA 杂交，通过与之杂交的 DNA 就可鉴定出该 RNA。特定基因的相对活性与连缀转录物和该基因 DNA 的杂交程度成比例。另外，可以改变连续反应的条件检测对转录物的影响。例如，添加 RNA 聚合酶抑制剂观察某一特定基因的转录是否被抑制，如果是的话，就可以鉴定出负责该基因转录的 RNA 聚合酶。

小结 核连缀转录是一种确定哪些基因在特定细胞中有活性的方法，做法是让这些基因的转录在细胞核分离后继续进行，通过与斑点膜上的已知 DNA 杂交鉴定特异性转录物。连缀转录分析也可用于检测核转录条件的变化对核转录的影响。

报告基因转录分析

另一种测定体内转录的方法是将一个替换性的报告基因置于特异启动子的控制下，然后检测该报告基因转录物的积累。设想要研究一个真核启动子的结构，一种方法是对含有该启动子的 DNA 区域做突变，然后将突变的 DNA 导入细胞，检测突变对启动子活性的影响。可以用 S1 图谱定位或引物延伸分析来检测，也可用报告基因替换天然基因，然后分析报告基因产物的活性。

为什么要用这种方法呢？主要原因是精心选择的报告基因的产物分析起来十分方便，比 S1 图谱定位或引物延伸方便得多。最常用的报告基因是 *LacZ*，其产物 β-半乳糖苷酶可通过发色底物 X-gal 被剪切时变成蓝色来检测。另一种广泛使用的报告基因是细菌中编码**氯霉素乙酰转移酶**（chloramphenicol acetyl transferase，**CAT**）的 *cat* 基因。大多数细菌的生长都受氯霉素（chloramphenicol，CAM）抑制，它能阻断蛋白质合成中的一个关键步骤（第 18 章）。有些细菌产生了一种逃避氯霉素抑制的方法，即对氯霉素乙酰化以阻断其活性，执行这种乙酰化作用的酶就是 CAT。但是真核生物对氯霉素不敏感，所以不需要 CAT。因此真核细胞中 CAT 活性的本底水平为零，这意味着如果将 *cat* 基因导入真核细胞并置于真核启动子之下，那么观察到的任何 CAT 活性都是由导入的该基因产生的。

怎样测定 *cat* 转染细胞中的 CAT 活性呢？最常

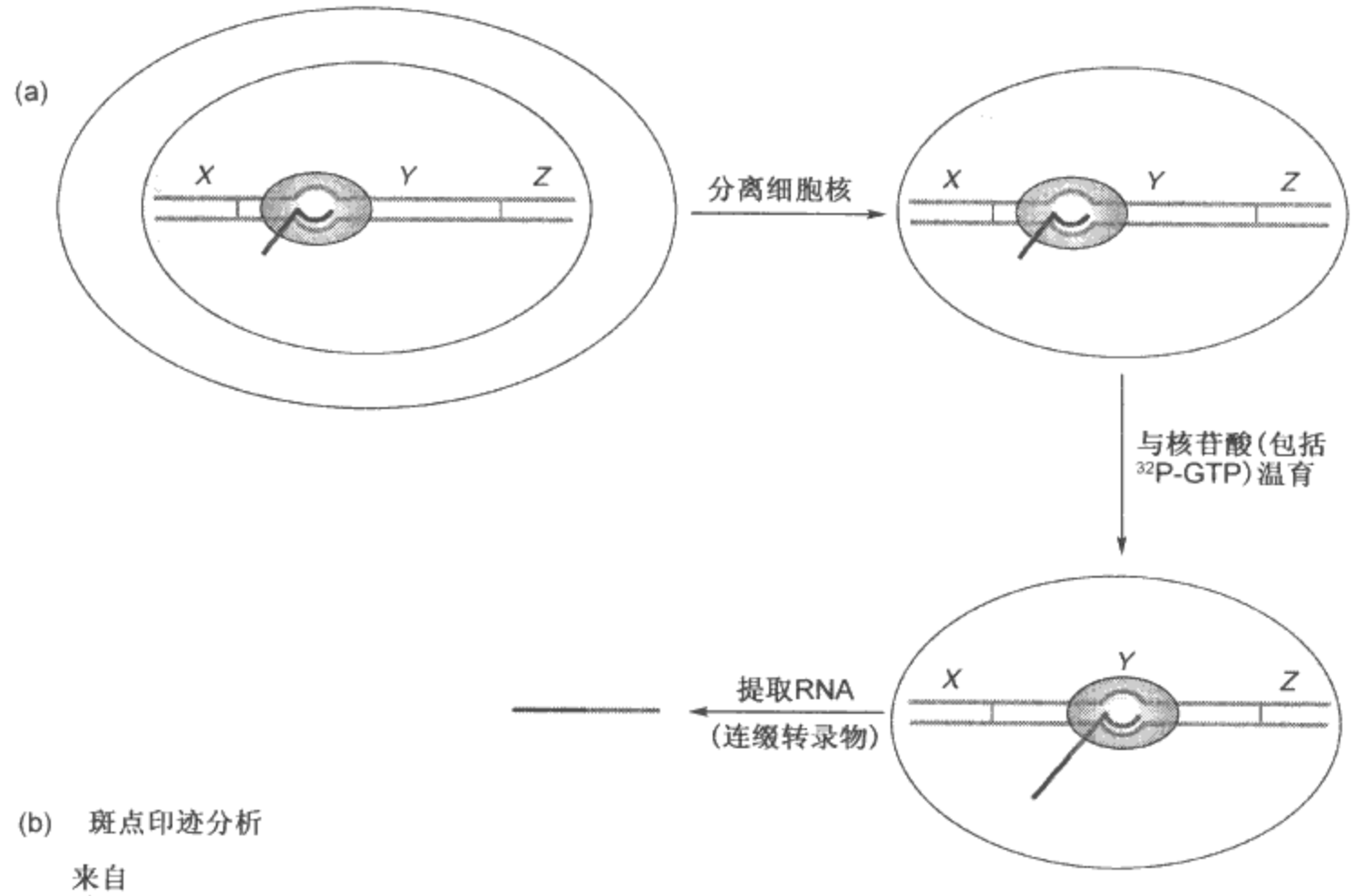


图 5.33 核连缀转录。（a）连缀转录反应。从正在进行 *Y* 基因转录的细胞开始。RNA 聚合酶（小椭圆）正合成 *Y* 基因的转录物（蓝色）。从细胞中分离细胞核并与核苷酸共孵育，使转录继续进行。在延续反应中还加入一种标记的核苷酸从而标记连缀转录物（红色）。最后提取标记的连缀转录物。（b）斑点杂交分析。将基因 *X*、*Y* 和 *Z* 的单链 DNA 分别点在硝化纤维素膜或其他合适的介质上与标记的连缀转录物杂交。因为 *Y* 基因是在连缀反应中转录的，其转录物被标记，因而 *Y* 基因的斑点也是标记的。*Y* 基因的转录活性越高，标记的信号就越强。由于 *X* 基因和 *Z* 基因没有活性，不会产生带标记的 *X* 和 *Z* 转录物，因此 *X* 和 *Z* 基因的斑点不被标记。

用的方法是将转染细胞的提取物与放射性标记的氯霉素和乙酰基供体（乙酰基辅酶 A）混合，然后使用薄层层析技术从乙酰化产物中分离氯霉素。乙酰化产物浓度越高说明细胞提取物中 CAT 的活性越高，进而说明启动子的活性也越高。图 5.34 显示这一过程。

（薄层层析利用粘贴到塑料板上的吸附性物质如硅胶形成薄层。将待分离的物质点在薄层板底部的加样点，然后将薄层板放入小室中，小室位于底部含有溶剂的浅槽中。当溶剂向上扩散时，分离物也向上移动，但其移动性取决于与吸附物和溶剂的相对亲和性。）

另一种标准的报告基因是从萤火虫分离的**萤光素酶**（luciferase）基因。萤光素酶与荧光素和 ATP 混合后将萤光素转化为一种能发光的化学发光物质，这就是萤火虫发光的秘密。发出的光很容易被 X 线胶片或闪烁计数仪检测，所以这种酶很适合做报告基因。

虽然基因产物来自包括转录和翻译的两步过程，但是在这个实验中我们假定报告基因的产物量是对转录速率（单位时间内 RNA 链起始的数量）以及启动子活性的测定。正常情况下，我们只改变启动子，那么假定翻译速率在不同的重组子 DNA 中不会改变是合理的。因为启动子位于编码区之外，其变化不会影响 mRNA 自身的结构，因而也不会影响翻译。不过，也可特意对基因待转录成 mRNA 的区域进行突变，然后用报告基因检测这些变化对翻译的影响。因此，根据基因突变的不同区域，报告基因也可以检测基因转录或翻译速率的变化。

小结 为了测定启动子活性，可以将启动子与

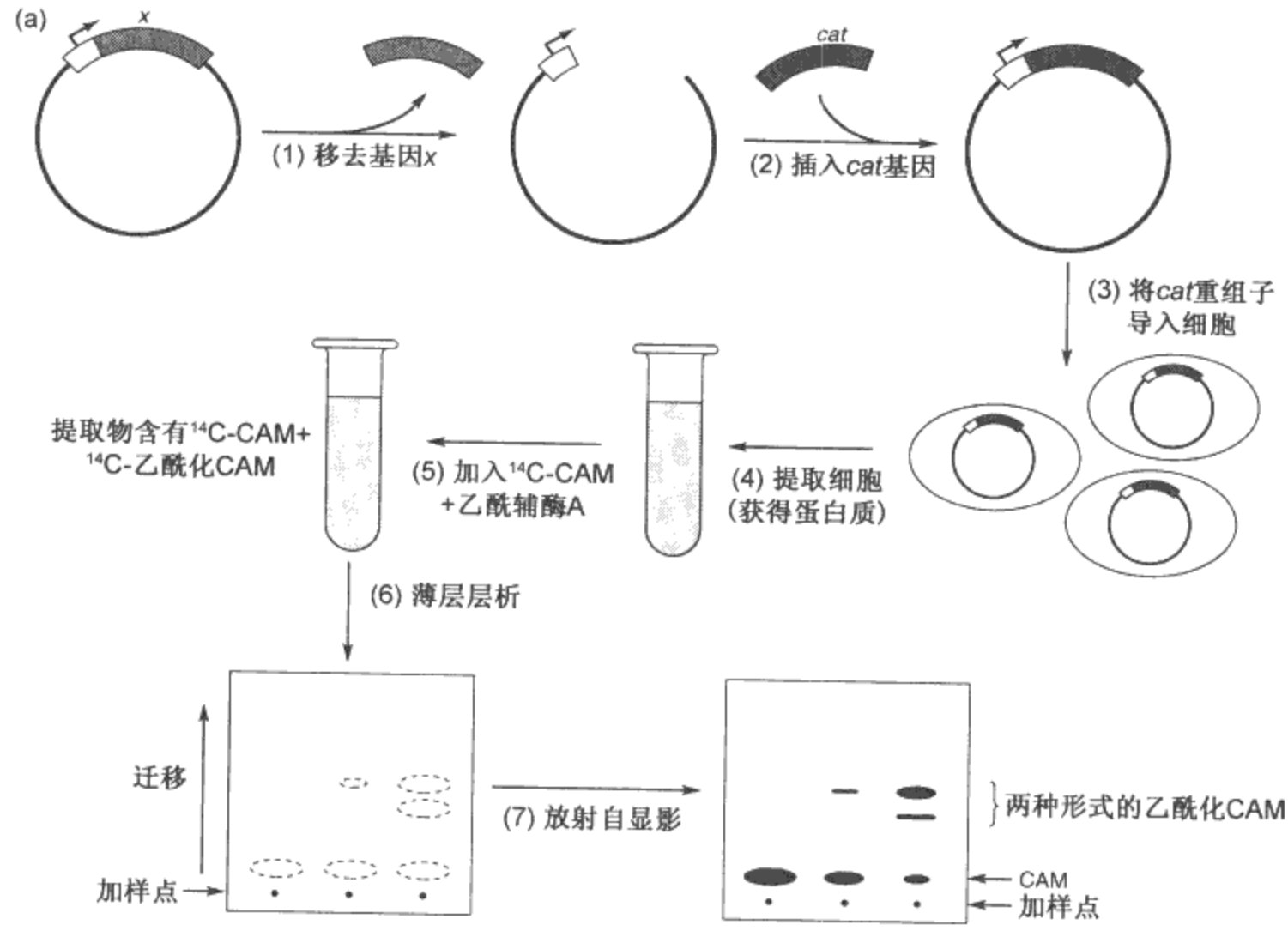


图 5.34　报告基因的使用。(a) 方法概要。步骤 (1)：从质粒开始，该质粒含有基因 x（蓝色）且该基因在自己的启动子（黄色）控制下，用限制酶切除基因 x 的编码区。步骤 (2)：将细菌的 *cat* 基因插入到 x 基因的启动子下。步骤 (3)：将构建的重组子转入真核细胞。步骤 (4)：一段时间后提取细胞内的可溶性蛋白。步骤 (5)：开始对 CAT 分析，加入 ^{14}C-CAM 和乙酰基供体乙酰辅酶 A。步骤 (6)：做薄层层析分离乙酰化和未乙酰化的氯霉素。步骤 (7)：最后，对薄层进行放射自显影观察 CAM 及其乙酰化的衍生物。这里可见 CAM 靠近放射自显影图的底部，两种乙酰化 CAM 因其高迁移率而靠近顶部。(b) 实际的实验结果。同样，原来的 CAM 靠近放射自显影的底部，两种乙酰化 CAM 靠近顶部。将薄层层析结果中有放射性的斑点刮下进行液体闪烁计数，得到 CAT 活性值列于图的底部（重复泳道的平均值）。泳道 1 是无细胞提取液的阴性对照。缩略语：CAM＝氯霉素；CAT＝氯霉素乙酰转移酶。（*Source*：(b) Qin, Liu, and Weaver. Studies on the control region of the p10 gene of the *Autographa californica* nuclear polyhedrosis virus. *J. General Virology* 70 (1989) f. 2, p. 1276. (Society for General Microbiology, Reading, England.)

一个报告基因连接，如编码 β-半乳糖苷酶、CAT、萤光素酶的基因，通过易于分析的报告基因产物来检测启动子活性。报告基因也可用来检测影响基因翻译的区域发生变化后翻译效率的变化。

体内蛋白质积累的检测

基因的活性也可以通过检测基因的终产物蛋白质的积累来测定。常用的两种方法是**免疫印迹法**

(Western blot) 和免疫沉淀法。免疫印迹法在本章前面部分已介绍过。

免疫沉淀法（immunopricipitation）从在细胞内标记蛋白质开始，做法是用标记的氨基酸（典型的是 [^{35}S] 甲硫氨酸）培养细胞。然后将标记的细胞匀浆，特殊标记的蛋白质结合到直接抗该蛋白质的特异性抗体或抗血清上。抗体-蛋白质复合物与第二抗体或蛋白 A 共沉淀，后者耦合了树脂磁珠可在低速离心下沉淀。沉淀的蛋白质电泳后通过放射自显影检测。注意，虽然沉淀中也有抗体和其他试剂，但由于未被标记所以检测不到。蛋白质条带上的标记越多，该蛋白质在体内积累的也越多。

小结 通过检测基因的蛋白质产物的积累可量化基因的表达，免疫印迹和免疫沉淀是完成这一任务的常用方法。

5.6 分析 DNA-蛋白质的相互作用

分子生物学的另一个永恒的主题是研究 DNA-蛋白质的相互作用。我们已经讨论过 RNA 聚合酶与启动子的相互作用，还会遇到更多这样的例子。因而需要一些方法来量化这些相互作用，检测 DNA 与给定蛋白质相互作用的确切部位。这里介绍两种检测蛋白质-DNA 结合的方法，给出 3 个确定 DNA 的哪些碱基与蛋白质相互作用的具体例子。

过滤结合

硝化纤维素（nitrocellulose）滤膜用于过滤除菌已经几十年了。分子生物学界传说是有人意外地发现 DNA 可以结合到硝化纤维素滤膜上，因为用这种方法制备 DNA 时发现 DNA 丢失了。这个故事的真假并不重要，重要的是硝化纤维素膜的确能在一定条件下结合 DNA。单链 DNA 容易结合到硝化纤维上，而双链 DNA 自身却不能。另外，蛋白质也能结合到硝化纤维上，如果蛋白质结合了双链 DNA，那么这个蛋白质-DNA 复合物就能结合在硝化纤维上，这是图 5.35 所示的分析方法的基础。

在图 5.35a 中，将标记的双链 DNA 倒入硝化纤维滤膜，分别对滤液（通过滤膜的物质）和滤膜吸附物中的标记量进行检测，结果显示所有被标记的物质都穿过滤膜进入滤液中，证明双链 DNA 不能结

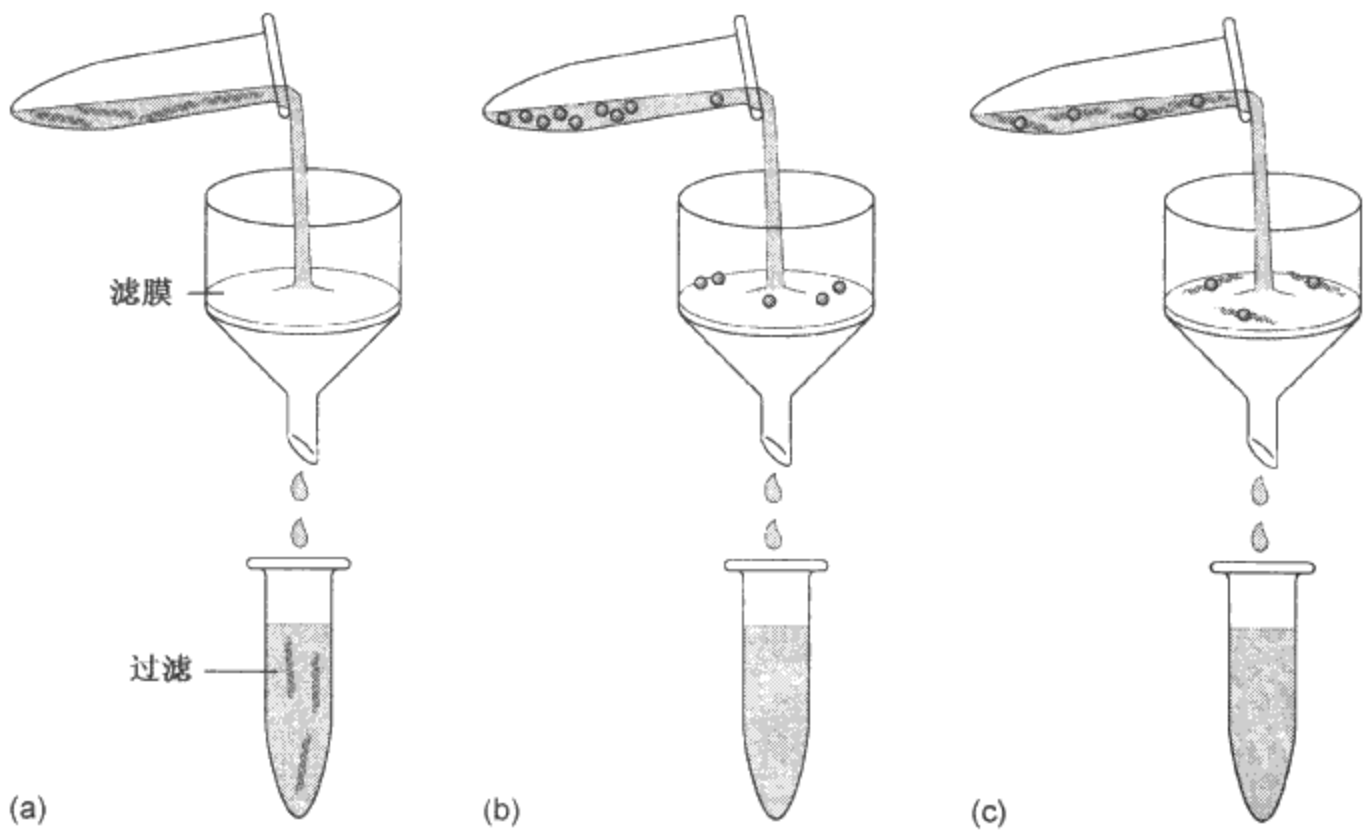


图 5.35 硝化纤维素滤膜结合分析。（a）双链 DNA。末端标记的双链 DNA（双螺旋形）通过硝化纤维素膜，用液体闪烁计数器检测膜上和滤液中的放射性。膜上没有放射性，说明双链 DNA 不能结合硝化纤维素膜。但是单链 DNA 可紧密结合在滤膜上。（b）蛋白质。标记蛋白质（小球）并流过硝化纤维素滤膜，蛋白质结合在滤膜上。（c）双链 DNA-蛋白质混合物。末端标记的双链 DNA 与未标记蛋白质（绿色）混合，形成 DNA-蛋白质复合物，然后流过硝化纤维素膜。由于与蛋白质的结合，标记的双链 DNA 结合在硝化纤维素膜上，由此为分析 DNA 与蛋白质的结合提供了一种方便的分析方法。

合在滤膜上。图 5.35b，过滤标记的蛋白溶液，所有的蛋白都结合在滤膜上，证明蛋白质自身与滤膜结合。图 5.35c，双链 DNA 又被标记，但这次它和能与之结合的蛋白混合在一起。由于蛋白质能与滤膜结合，所以蛋白质-DNA 复合物也能结合，因而是在滤膜上而不是滤液中有放射性。因此，滤膜结合是一种直接检测蛋白质-DNA 相互作用的方法。

小结 滤膜结合作为检测蛋白质-DNA 相互作用的方法基于双链 DNA 自身不能结合硝化纤维素滤膜或类似介质，而蛋白质-DNA 复合物可以结合这一性质。因此，可以标记双链 DNA，将其与某种蛋白质混合，通过检测滞留在滤膜上的标记量分析 DNA 与蛋白质之间的相互作用。

凝胶阻滞

另一种检测 DNA-蛋白质相互作用的方法依据小分子 DNA 在凝胶电泳中的迁移率快于它与蛋白质结合后的迁移率这一性质。因此，将一个短的双链 DNA 片段标记后与某种蛋白质混合并进行凝胶电泳，然后对电泳凝胶进行放射自显影检测标记物。图 5.36 显示的是 3 个不同样本的电泳迁移率。泳道 1 是裸 DNA，因为体积小，所以迁移率很高。在本章前面讲过，DNA 电泳图的顶部习惯性地作为起始位置，因此高迁移率的 DNA 出现在凝胶的底部附近。泳道 2 是结合了蛋白质的 DNA，其迁移率大大降低。这就是该技术名称：**凝胶阻滞分析**（gel mobility shift assay）或**电泳阻滞分析**（electrophoretic mobility shift assay，EMSA）的来源。泳道 3 是同样的 DNA 结合了两个蛋白质的电泳结果，因为结合在 DNA 上的蛋白质分子质量更大，迁移率进一步降低，这叫做**超阻滞**（supershift）。该蛋白可能是另一种 DNA 结合蛋白也可能是结合在第一个蛋白质上的第二个蛋白质，甚至还可能是一种与第一个蛋白质特异性结合的抗体。

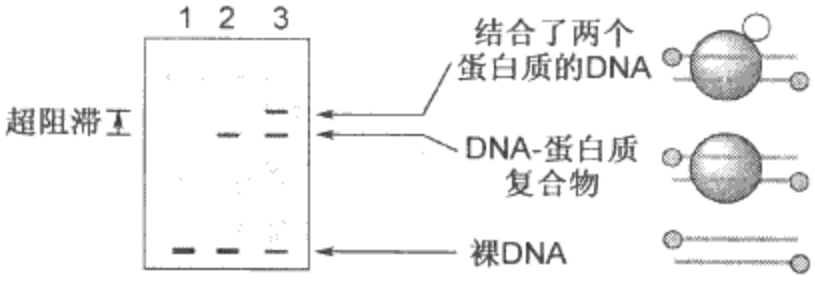


图 5.36 凝胶阻滞实验。对纯标记的 DNA 或 DNA-蛋白质复合物进行凝胶电泳，然后放射自显影检测 DNA 及其复合物。泳道 1 显示裸 DNA 的高迁移率。泳道 2 显示蛋白质结合 DNA 后迁移的阻滞。泳道 3 显示 DNA-蛋白质复合物再结合第二个蛋白质后的超阻滞现象。DNA 两端的小圆点代表末端标记物。

小结 凝胶迁移阻滞分析通过将小分子的 DNA 与蛋白质结合降低 DNA 分子的电泳迁移率来分析 DNA 与蛋白质的相互作用。

DNA 酶足迹

足迹法是检测蛋白质与 DNA 相互作用的方法，可确定 DNA 上的靶位点，甚至可确定参与蛋白质结合的碱基。有多种足迹法，较通用的 3 种是：DNase、硫酸二甲酯（DMS）和**羟基自由基足迹法**（hydroxyl radical footpringting）。**DNase 足迹法**（DNase footpringting）（图 5.37）基于蛋白质与 DNA 结合后覆盖了结合位点使其免受 DNase 降解的原理。在这个意义上，蛋白质将其“足迹”就留在了 DNA 上。足迹实验的第一步是对 DNA 进行末端标记。双链的任意一条链都可被标记，但每次实验只能标记一条。接着是蛋白质（图中椭圆部分）与 DNA 结合，然后用 DNase I 在温和条件下（很少量的 DNase I）处理 DNA-蛋白质复合物，使每个 DNA 分子平均仅被切割一次。接下来，将蛋白质从 DNA 上除去，分离出 DNA 链，所得 DNA 片段在高分辨率的聚丙烯酰胺凝胶上电泳，在旁边泳道加入 DNA 标准分子（图上未给出）。当然，DNA 的另一端也会产生片段，但是没有标记因此检测不到。电泳时，通常包括一个无蛋白质的 DNA 作对照，同时还设置一个以上的蛋白质浓度，足迹区域 DNA 条带的逐渐消失说明 DNA 受保护的程度依赖于蛋白质浓度。足迹代表了被蛋白质保护的 DNA 区域，因而也指示了蛋白质结合的部位。

DMS 足迹法和其他足迹法

DNase 足迹法提供了研究蛋白质在 DNA 上定位的很好思路，然而 DNase 是一种大分子，对于探究结合位点的详细情况来说显得过于钝拙，即在蛋白质和 DNA 相互作用的区域可能会出现缺口，但 DNaseI 不能进入而无法检测。此外，DNA 结合蛋白经常干扰结合区域的 DNA，使双螺旋结构扭曲，尽管这些干扰作用是有意义的，但是一般不能用 DNase 足迹法检测，因为蛋白质使 DNase I 不能接近。而更细致的足迹法需要一种更小的分子以便进入 DNA-蛋白质复合物的隐蔽处和缝隙里，从而揭示这种相互作用的细微之处。**硫酸二甲酯**（dimethyl sulfate，DMS）可很好地完成这一任务。

图 5.38 显示 DMS 足迹法，与 DNase 足迹法一样，以末端标记的 DNA 与蛋白质结合开始。然后，复合物经 DMS 温和甲基化，使每个 DNA 分子平均只发生一次甲基化。接着去除蛋白质，DNA 用六氢吡啶处理，去除甲基化的嘌呤形成脱嘌呤位点（无

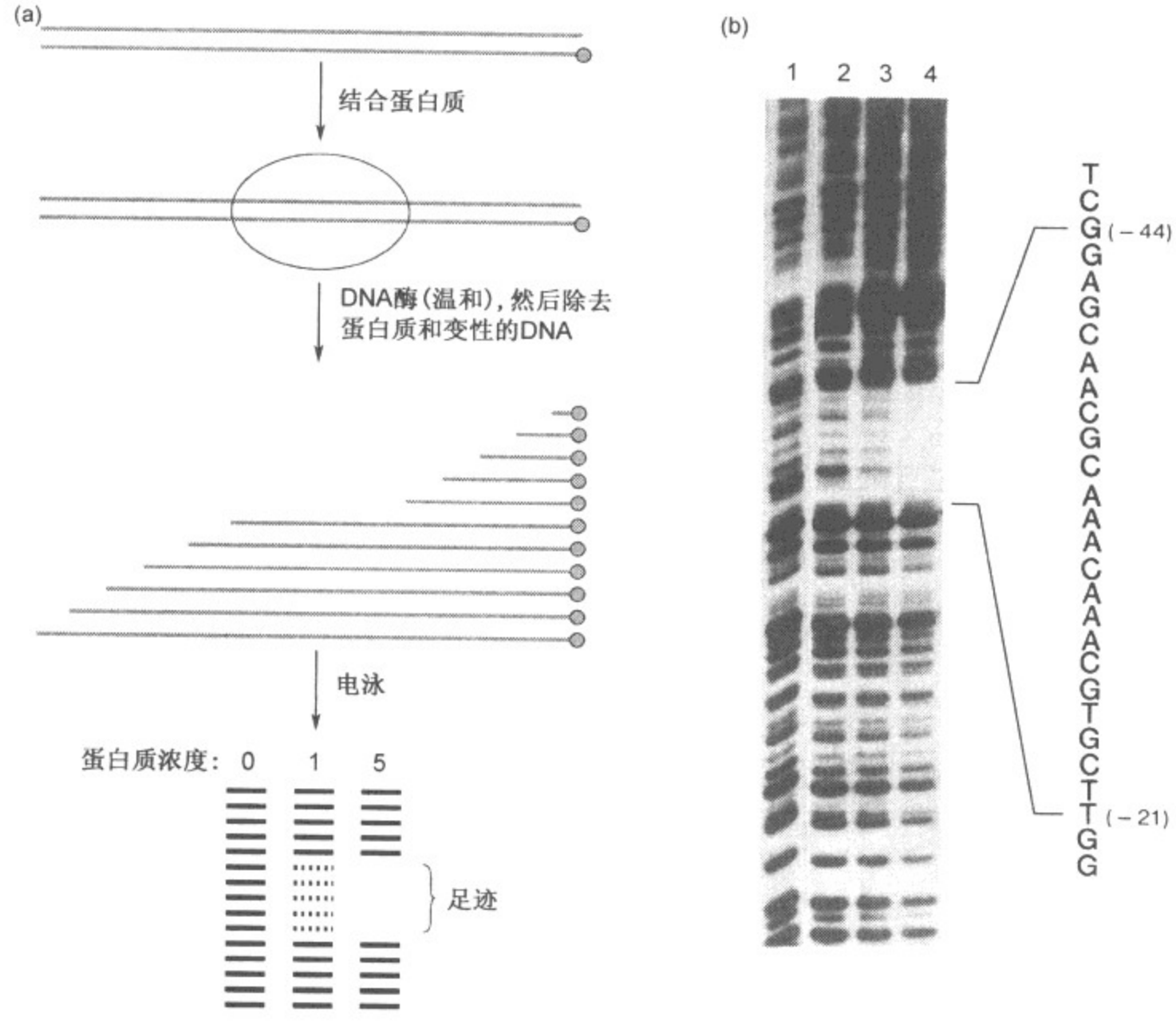


图 5.37　DNA 酶足迹实验。（a）方法概要。以一端带标记（小圆点）的双链 DNA 开始。接着将一个蛋白质与其结合。然后在温和条件下用 DNase I 处理 DNA-蛋白质复合物，以便对每分子 DNA 约引入一个切口。然后除去蛋白质，变性 DNA，产生末端标记的片段如图中部所示。注意，除蛋白质结合区域外，DNase 以规则的间隔切割 DNA。最后，电泳标记的片段并放射自显影检测。3 个泳道分别代表结合了 0、1、5 个单位蛋白质的 DNA。没有蛋白质的泳道显示出片段的正常梯形，有 1 单位蛋白质的泳道显示部分保护，有 5 单位蛋白质的泳道显示中部完全被保护。被保护的区域就叫做足迹，显示的是蛋白质结合 DNA 的部位。通常还要在平行泳道中对相同的 DNA 进行测序反应，指出蛋白质结合的位点。（b）实际的实验结果。泳道 1～4 分别是结合了 0pmol、10pmol、18pmol、90pmol 蛋白质的 DNA（1pmol＝10^{-12}mol）。先前已经通过标准的双脱氧测序法获得 DNA 的序列。（*Source*：（b）Ho et al., Bacteriophage lambda protein *c*II binds promoters on the opposite face of the DNA helix from RNA polymerase. *Nature* 304（25 Aug 1983）p. 705, f. 3, © Macmillan Magazines Ltd.）

碱基的脱氧核糖核苷酸），并在这些脱嘌呤位点处断裂 DNA。

最后电泳 DNA 片段，凝胶放射自显影检测被标记的 DNA 条带。每个条带的两端接着因甲基化而未被蛋白质保护的碱基。在这个例子中随着更多的蛋白质加入，有 3 个条带逐渐消失，在高浓度下，实际上只有一个条带是显著的。这说明与蛋白质的结合干扰了 DNA 的双螺旋结构从而导致相应条带的碱基对甲基化更敏感。

除 DNase I 和 DMS 外，其他一些试剂也常用于 DNA-蛋白质复合物足迹法，打断蛋白质保护区域之外的 DNA 分子。例如，含有铜或铁有机金属的复合物通过产生**羟基自由基**（hydroxyl radical）来攻击并打断 DNA 链。

小结　足迹法是寻找 DNA 结合蛋白的靶 DNA 序列或结合位点的一种方法。DNase 足迹法的操作步骤是：先将蛋白质与其靶 DNA 结合，然后用 DNase 处理 DNA-蛋白质复合物。所得 DNA 片段进行电泳时，在蛋白质结合位点呈现为

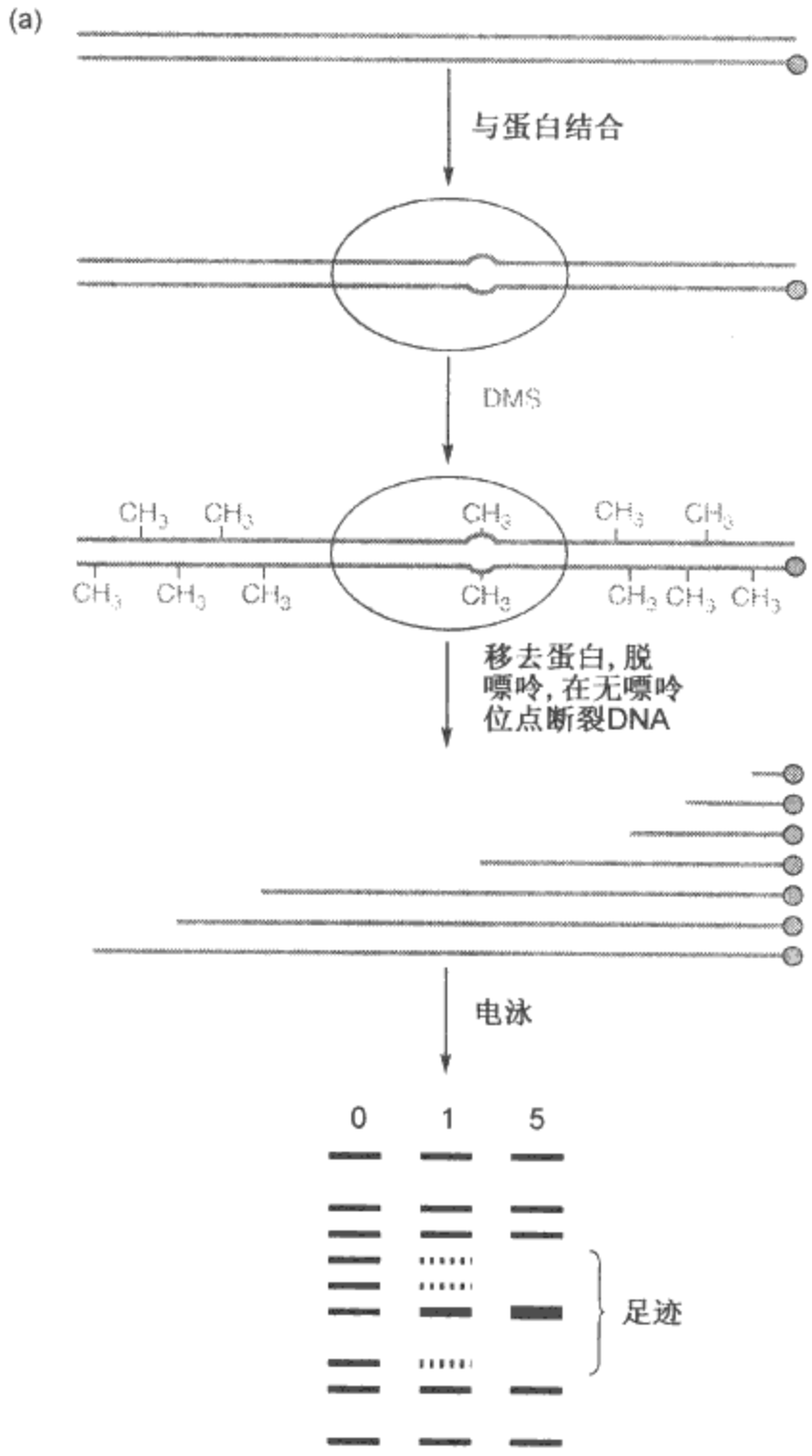


图 5.38　硫酸二甲酯（DMS）足迹法。（a）方法概要。与 DNA 酶足迹法相同，以末端标记的 DNA 开始，然后与蛋白质（椭圆）结合。在本例中，蛋白质与 DNA 结合后引起双链 DNA 的某个区域融解，形成一些“小泡”。用 DMS 甲基化 DNA，使 DNA 的某些碱基带上甲基基团（CH_3）。甲基化在温和条件下进行，使每个 DNA 分子平均仅有一个碱基甲基化（这里为方便起见将所有 7 处甲基化都标在一条链上）。然后用六氢吡啶从 DNA 上除去甲基化的嘌呤，接着在这些脱嘌呤位点处断裂 DNA，由此产生图中间所示的带标记的 DNA 片段。片段经电泳后放射自显影，结果显示在图底部。有 3 个位点被蛋白质保护未被甲基化，但有一个位点对甲基化却更敏感了（较黑的条带），因为蛋白质结合后在此位点发生了解旋。（b）实际的电泳结果。泳道 1 和泳道 4 无蛋白质，泳道 2 和泳道 3 结合该 DNA 区域的蛋白质浓度逐渐增加。括号内是一个显著的足迹区域。星号指蛋白质结合后产生的甲基化敏感区域。（*Source*：（b）Learned et al.，Human rRNA transcription is modulated by the coordinate binding of two factors to an upstream control element. *Cell* 45（20 June 1986）p. 849，f. 2a. Reprinted by permission of Elsevier Science.）

缺口，或“足迹”，其形态为蛋白质保护 DNA 免受降解的方式。硫酸二甲酯（DMS）足迹法遵循类似的原理除了用 DNA 甲基化试剂 DMS 代替 DNase 攻击 DNA-蛋白质复合物外。DNA 在甲基化位点被打断。标记的 DNA 片段在电泳上显示的未甲基化（或超甲基化）位点表明了 DNA 与蛋白质结合的部位。羟基自由基足迹法用含有铜或铁有机金属的复合物产生羟基自由基打断 DNA 链。

5.7 寻找与其他分子相互作用的RNA序列

SELEX

SELEX[**指数级富集配体系统进化技术**（systematic evolution of ligands by exponential enrichment）]最初研发出来是用于寻找与特殊分子结合的短RNA序列［适配子（aptamer）］的一种方法。图5.39显示了经典SELEX的过程。以PCR扩增合成DNA的混合物开始，这些DNA具有不变的两端区域（红色）和随机的中间区域（蓝色），该随机区域可以编码超过10^{15}的不同RNA序列。第一步，用噬菌体T7 RNA聚合酶体外转录这些DNA，T7RNA聚合酶识别混合物中每个DNA不变区上游的T7启动子。接下来，用树脂结合了靶分子的亲和柱层析筛选适配子。选出的RNA结合到树脂上然后用含有靶分子的溶液洗脱。用DNA常规区末端特异性的引物，将选出的RNA反转录成双链DNA，再次进行PCR。

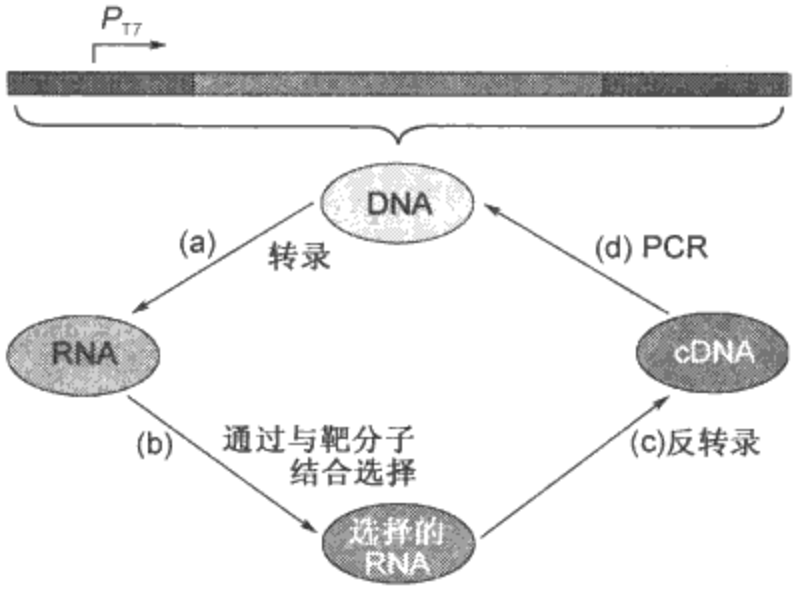


图5.39 SELEX。以一个大DNA的集合（顶部）开始，其序列是随机的（中间部分）两侧是不变序列。（a）将DNA库转录成RNA库，其中含有两侧是不变序列的随机序列；（b）通过带有靶分子的亲和层析选择适配子；（c）反转录RNA产生cDNA库；（d）PCR扩增cDNA，引物与DNA两端不变区互补，将这个循环重复几次富集库中的适配子。

一轮SELEX产生的一大批分子只是部分富集了适配子，所以这一过程要重复多次才能产生高度富集的适配子。SELEX已经被广泛地应用于寻找与蛋白质结合的RNA序列，尤其是在天文数字般巨大的起始RNA序列中寻找几个适配子时极为有效。

功能性SELEX

功能性SELEX（functional SELEX） 在起始序列的“大海”中捞几个“针”（RNA序列）方面与经典的SELEX很相似，但不是寻找与其他分子结合的适配子，而是寻找具有一定功能的RNA序列。仅仅是简单的结合，筛选还是容易的，只需亲和层析即可。但是基于功能的筛选就复杂多了，在设计筛选步骤中需要有创造性。例如，最早的功能性SELEX程序是检测核酶（具有酶活性的RNA），这种核酶活性改变了RNA自身使其能够被扩增。一个简单的例子是核酶可以给自己的末端添加核苷酸，这一活性使研究者可以给核酶提供序列已知的寡聚核苷酸，而核酶又可将这个标签加到自身。一旦加上标签，核酶就成为可用PCR引物与标签互补进行扩增的底物了。

随机RNA序列库也许不含任何高活性的RNA，但是这一问题可通过在突变条件下的扩增而克服，由此产生多种有活性的变异体，其中有些分子的活性可能比其原来的要高，几轮筛选和突变之后就可以产生具有很强酶活性的RNA。

小结 SELEX是一种寻找与其他分子，包括蛋白质相互作用的RNA序列的方法。通过亲和层析筛选与靶分子作用的RNA，然后将其转换成双链DNA并进行PCR扩增。这个程序经过几轮后，与靶分子结合的RNA序列就被极大地富集了。功能性SELEX是这一方法的衍生，其中期望的功能是改变RNA使其能够被扩增。如果期望的功能是酶学上的，便可在扩增步骤中引入突变以产生具有更高活性的变异体。

5.8 基因敲除

在第5章中我们所讨论的大部分技术都是关于探索基因的结构和活性的，然而这些技术经常遗留一个有关基因功能的大问题：在生物体的生命活动中这些基因起什么作用？通常，对这个问题最好的回答是观察生物体内某个基因突变后的情况。我们现在有了靶向中断生物体内基因的方法。例如，可以通过插入外源片段而中断小鼠的基因，并把这种小鼠叫做**敲除小鼠**（knockout mice）。

图5.40解释了一种构建基因敲除小鼠的方法。首先克隆一段DNA，其中含有待敲除的小鼠基因，然后用具有新霉素抗性的基因插入，中断这个靶基因，在克隆基因的其他部位（靶基因之外）引入一个胸苷激酶基因（tk^-）。这两个额外基因将用于随后

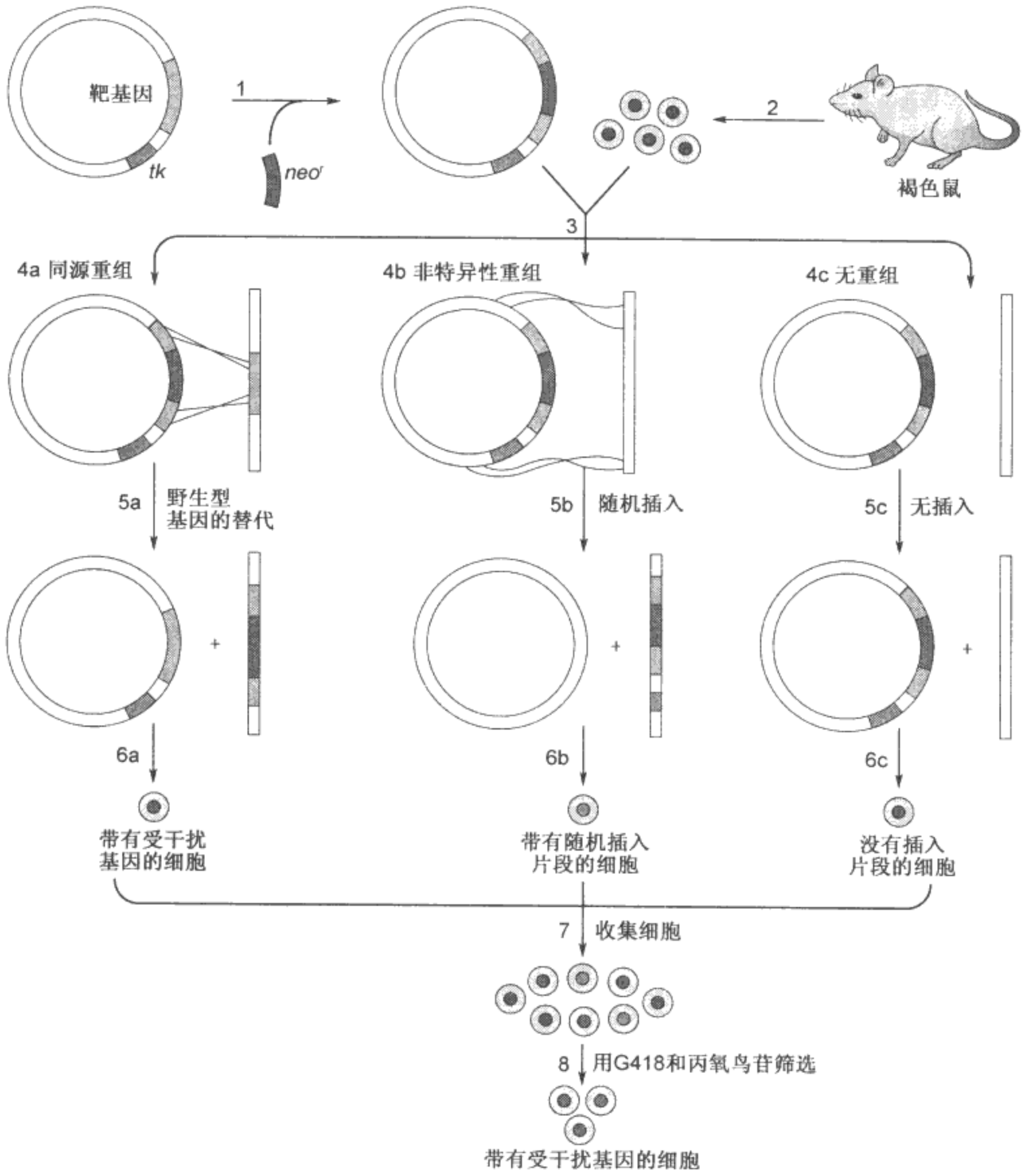


图 5.40　构造基因敲除小鼠：阶段 1，构建带有中断基因的干细胞。1. 以含有待敲除基因（靶基因）和胸苷激酶基因（tk^+）的质粒开始。将新霉素抗性基因（红色）插入靶基因内部中断靶基因。2. 收集褐色小鼠的胚胎干细胞。3. 用含有打断了靶基因的质粒转染胚胎干细胞。4 和 5. 转染后产生 3 种结果：4a. 质粒上被中断的靶基因与同源的野生型靶基因发生同源重组，使染色体的野生型靶基因被质粒的中断靶基因取代（5a）。4b. 与细胞基因组的非同源序列发生非特异性重组，使中断靶基因和 *tk* 基因随机插入细胞的基因组（5b）。4c. 未发生重组，中断靶基因没有与基因组整合（5c）。6. 由这 3 种情况产生的细胞分别用不同颜色的表示：同源重组产生带有中断靶基因（6a）的细胞（红色）；非同源重组产生带有随机插入的中断靶基因与 tk^+ 基因的细胞（蓝色）；非重组产生没有中断靶基因整合的细胞（棕色）。7. 收集转化的细胞含有全部 3 种细胞（红、蓝、棕）。8. 在含有新霉素类似物 G418 和丙氧鸟苷的培养基中培养细胞。G418 可杀死无新霉素抗性的细胞，即未发生重组的细胞（棕色）。丙氧鸟苷杀死含 *tk* 基因的细胞，即发生非特异性重组的细胞（蓝色），最后只留下发生同源重组的细胞（红色），因而含有中断靶基因。

剔除未发生定向敲除的克隆。

接下来，将构建的小鼠 DNA 与褐色小鼠的胚胎干细胞混合。中断的基因通过某种途径进入少许胚胎干细胞的核内，与细胞中同源的野生型靶基因之间发生同源重组，使被中断的基因进入小鼠基因组，并除去胸苷激酶基因（*tk*）。遗憾的是发生这种重组的概率非常小，许多干细胞不会发生重组因此也不会受到中断的靶基因的干扰。还有一些干细胞会发生非特异性重组，被中断的基因随机插入到基因组中而没有替换相应的野生型基因。

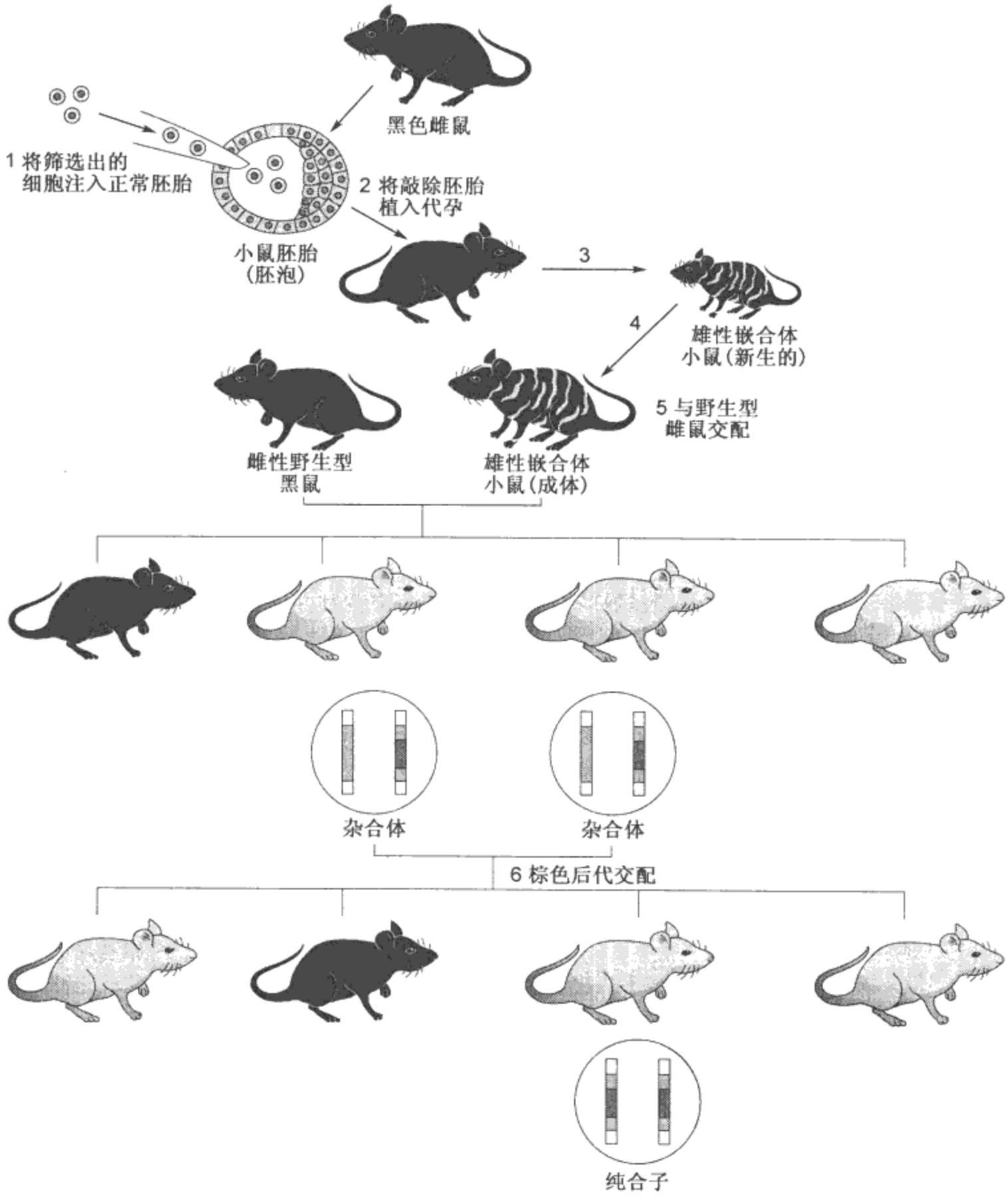


图 5.41　构造基因敲除小鼠：阶段 2，将中断基因置入动物体内。1. 将含有中断基因（图 5.40）的细胞注射到黑鼠的囊胚期胚泡中。2. 将这个混合胚胎植入代孕母鼠的子宫中。3. 代孕母鼠产下嵌合体小鼠，其黑色和棕色皮毛可以鉴定（回想一下，这些改变的细胞来自棕色鼠，被置于黑色鼠的胚胎中）。4. 让嵌合体小鼠（雄性）成熟。5. 将其与野生型雌性黑鼠交配，丢弃来源于野生型胚胎的黑色后代，只有棕色小鼠来自转基因移植的细胞。6. 选择一对均含有中断基因（通过 Southern 印迹检测）的雌雄棕色小鼠并让它们交配。再用 Southern 印记检测棕色后代的 DNA。这一次找到了中断基因的纯合体，即基因敲除小鼠。现在观察这只小鼠以确定靶基因敲除所造成的影响。

那么如何剔除未发生同源重组的细胞呢？我们先前引入的额外基因在这个环节上便可发挥作用了。未发生重组的细胞没有新霉素抗性基因，因此在含有新霉素衍生物 G418 的培养基上培养细胞可以将其排除。而发生了非特异性重组的细胞，*tk* 基因随干扰基因一同进入细胞的基因组中，用一种能杀死 tk^+ 细胞的药物 gangcyclovir 杀死这些细胞（我们用的干细胞是 tk^-）。用这两种药物处理后，剩下的就是发生了同源重组的工程细胞，而且是含有中断基因的杂合子细胞。

下一个任务是将这些中断基因导入小鼠体内（图 5.41），做法是将得到的工程干细胞注射到最终将发育成黑鼠的胚泡中，再将这个改变的胚胎植入代孕母鼠子宫。母鼠产下嵌合体小鼠。通过小鼠的毛色可辨认嵌合体小鼠，其黑色条纹来自原来的黑鼠胚胎，而褐色条纹来自移植的工程细胞。

为了获得真正的杂合子小鼠而不是嵌合体，可等到嵌合体小鼠成熟后与黑鼠交配。由于褐色（鼠灰色）是显性性状，所以子代中自然有褐色小鼠。实际上，所有由工程干细胞的配子发育成的子代小鼠都是褐色的。由于工程干细胞是敲除基因的杂合子，所以只有一半的褐色小鼠携带有中断基因。Southern 印迹显示在我们的实验中有两只褐色小鼠携带中断基因。将它们交配后在子代小鼠中检测 DNA 寻找基因敲除的纯合子小鼠。在这个例子中，交配后得到一只基因敲除的纯合子小鼠。现在的任务是观察其表型。表型经常并不显著（你能看出吗?），但是无论显著与否都是非常有指导意义的。

有时基因敲除是致死性的，受影响的小鼠胎儿在出生前就已死亡。还有一些基因敲除表现出中间效应。例如，肿瘤抑制基因 *p53*，缺失此基因的人极易患上某些癌症。*p53* 基因敲除小鼠能够正常发育但是在幼年期会受肿瘤之苦。

小结 为探寻某个基因的功能，分子生物学家可实施对小鼠相应基因的靶向敲除，然后观察突变对基因敲除小鼠的影响。

总结

蛋白质与核酸纯化的方法在分子生物学中是至关重要的。通过凝胶电泳可以分离各种大小不同的 DNA、RNA 和蛋白质。核酸电泳最常用的凝胶是琼脂糖，蛋白质电泳一般选用聚丙烯酰胺凝胶。SDS 聚丙烯酰胺凝胶电泳（SDS-PAGE）可以通过分子的大小分离多肽。利用双向电泳分离多肽可获得高分辨率的效果，第一向为等电聚焦和第二向为 SDS-PAGE。

离子交换层析技术根据所带的电荷分离蛋白质等物质。带正电荷的树脂（如 DEAE-葡聚糖）用于阴离子交换层析，带负电荷的树脂（如磷酸纤维素）用于阳离子交换层析。凝胶过滤层析利用填满了多孔树脂的柱子使小分子物质进入孔内而将大分子排除在孔外，因此小分子物质流速缓慢而大分子物质相对快速地通过柱子。亲和层析是一种有效的分离技术，利用了对目的分子有很强专一性的亲和试剂。目的分子结合到耦合了亲和试剂的柱子上，而所有或大多数其他分子能不结合，直接流过柱子，然后目的分子可以用打断专一性结合的溶液洗脱下来。

分子生物学实验中检测微量物质通常需要标记的示踪剂。如果示踪剂是放射性的则可通过 X 线胶片或磷屏成像技术进行放射自显影检测，或用液体闪烁计数仪检测。现在也有非常灵敏的非放射性标记的示踪剂，它们能产生光（化学发光）或有色斑点。

标记的 DNA 或 RNA 探针能与 Southern 印迹膜上相同的或非常相似的序列杂交。现代 DNA 分型技术使用 Southern 印迹和一组 DNA 探针，检测动物个体（包括人类）中的变异位点。作为一种取证工具，DNA 指纹可用于亲子鉴定、罪犯鉴别或排除嫌疑人。

标记的探针可与整个染色体杂交以便对基因或其他特异性 DNA 序列进行定位，这种杂交的方式称为原位杂交，如果探针是荧光标记的，则该技术称为荧光原位杂交（FISH）。复合物中的蛋白质可以用免疫印迹（或 Western 印迹）探测和定量。将蛋白质电泳后印迹到膜上，然后用专一性抗体来探测，而专一性抗体可被标记的第二抗体或蛋白 A 检测。

Sanger DNA 测序法利用双脱氧核苷酸终止 DNA 合成，产生一系列 DNA 片段，其大小可通过电泳检测。每个片段的最后一个碱基是已知的，因为用于终止每种反应的双脱氧核苷酸是已知的。因此将片段按大小排列，每个片段比下一个多一个碱基（碱基已知），就能获得 DNA 的碱基序列。

物理图谱描述 DNA 分子的物理“路标”，如限制酶位点的空间排列顺序。通过 Southern 印迹法可完美地绘制限制性图谱，即将一种酶切片段进行 Southern 印迹，然后与另一种限制性内切核酸酶产生的被标记的酶切片段进行杂交，这样就能知道两个不同限制酶酶切片段之间的重叠情况。

利用克隆的基因可以很方便地通过定点诱变引入突变，从而改变蛋白质产物的氨基酸序列。突变的 DNA 可以通过双链 DNA、两条互补的突变引物，通过 PCR 的方法产生。野生型 DNA 可用 *Dpn*I 酶切去除，这样得到的克隆都是由突变的而不是野生

型 DNA 转化的。

Northern 印迹与 Southern 印迹类似，但它电泳分离的是 RNA 而不是 DNA。印迹膜上的 RNA 可通过与标记探针杂交进行检测。条带亮度表示每个条带中特定 RNA 的相对含量，条带的位置说明该特定 RNA 的大小。

S1 图谱定位中，标记的 DNA 探针用于检测转录物 3′端或 5′端。探针与转录物杂交使探针的被杂交部分免受 S1 核酸酶的消化，该酶特异性地降解单链多核苷酸。根据被转录物保护的探针部分的长度就能定位转录物的末端，它是相对于探针一个已知末端的定位。由于受转录物保护的探针量与转录物的浓度成正比，因此 S1 图谱法也可作为一种定量方法使用。RNA 酶图谱源自 S1 图谱法，不同的是它用 RNA 探针和 RNA 酶取代了 DNA 探针和 S1 核酸酶。

用引物延伸法可以定位转录物的 5′端，做法是用一个寡核苷酸引物与目的 RNA 杂交，用反转录酶向转录物的 5′端延伸引物，凝胶电泳确定反转录物的大小，由此得到的信号强度可作为转录物含量的测定。

截断转录是一种检测体外转录效率和精确性的方法。将基因从中部截断，并用标记的核苷酸进行体外转录。RNA 聚合酶脱离末端，释放不完整的转录物。根据截断转录物的大小可定位转录的起始位点，转录物的量反映了转录的效率。无 G 盒转录也能产生一段大小可预测的被截短的转录产物，但它是将一个无 G 盒置于启动子下游并在缺少 GTP 条件下转录形成的。核连缀转录是一种确定哪些基因在特定细胞中有活性的方法，做法是让这些基因在分离的细胞核里继续转录。特定的转录物可通过与已知 DNA 的斑点杂交来确定。利用连缀转录分析也可测定分析不同的条件对细胞核转录的影响。

为了测定启动子的活性，可以将启动子和报告基因连接，如 β-半乳糖苷酶、CAT 或萤光素酶，通过易于分析的报告基因产物来显示启动子的活性。报告基因也可用于检测影响基因翻译的区域发生变化后翻译效率的变化。基因表达也可以量化，做法是通过免疫印迹或免疫沉淀检测蛋白质产物的积累。

滤膜结合实验是一种测量 DNA 与蛋白质相互作用的方法，其原理是双链 DNA 自身不能与消化纤维素膜或其他类似介质结合，而 DNA-蛋白质复合物却可以与膜结合。因此，将标记的双链 DNA 与蛋白质混合后，通过检测膜上残留的标记物的量来分析 DNA 与蛋白质的结合情况。凝胶阻滞实验根据小分子 DNA 与蛋白质结合后能降低 DNA 的电泳迁移率来分析 DNA 与蛋白质之间的相互作用。

足迹法是寻找 DNA 结合蛋白的靶 DNA 序列或结合位点的一种方法。DNase 足迹法的操作步骤是：先将蛋白与其靶 DNA 结合，然后用 DNase 处理 DNA-蛋白质复合物，所得 DNA 片段进行电泳时，在蛋白质结合位点呈现为缺口或“足迹”，其带型为蛋白质保护 DNA 免受降解的方式。除了用 DNA 甲基化试剂硫酸二甲酯（DMS）代替 DNase 攻击 DNA-蛋白质复合物外，DMS 足迹法均遵循类似的原理。电泳显示未甲基化（或超甲基化）位点，表明 DNA 与蛋白质结合的部位。羟基自由基足迹法用有机金属复合物产生羟基自由基打断 DNA 链。

SELEX 是一种寻找与其他分子，包括蛋白质相互作用的 RNA 序列的方法。通过亲和层析筛选与靶分子作用的 RNA，然后将其转换成双链 DNA 并进行 PCR 扩增。这个程序经过几轮后，与靶分子结合的 RNA 序列就被极大地富集了。功能性 SELEX 是这一方法的衍生，其中期望的功能可改变 RNA 使其能够被扩增。如果期望的功能是酶活性的，可在扩增步骤引入突变以产生具有更高活性的变异体。

为探寻基因的功能，可以通过定向敲除小鼠中相应的基因，然后观察这种突变对基因敲除小鼠的影响。

复习题

1. 绘图说明 DNA 电泳的原理，大致表示出 150bp、600bp 和 1200bp DNA 片段相对的电泳迁移率。

2. 什么是 SDS？它在 SDS-PAGE 中的作用是什么？

3. 比较 SDS-PAGE 与蛋白质现代双向凝胶电泳的异同。

4. 描述离子交换层析的原理。用曲线图说明用这种方法分离 3 种不同的蛋白质。

5. 描述凝胶过滤层析的原理。用曲线图说明用这种方法分离 3 种不同的蛋白质，指出图中最大与最小的蛋白质。

6. 比较放射自显影与磷屏成像原理的异同，哪种方法能提供更量化的信息。

7. 描述用非放射性方法检测电泳凝胶中的特定核酸片段。

8. 图解 Southern 印迹和探针检测目的 DNA 的过程，并比较与 Northern 印迹法的异同。

9. 简述以微卫星为探针的 DNA 指纹技术，并将该方法与现代法医 DNA 分型技术中利用探针检测单一可变 DNA 位点的方法做一比较。

10. 从 Northern 印迹膜可以获得哪类信息？

11. 描述荧光原位杂交（FISH），什么时候使用

这种方法而不用 Southern 印迹?

12. 绘制一个假想的 Sanger 测序法的放射自显影图，并给出正确的 DNA 序列。

13. 说明如何将手动的 DNA 测序法自动化。

14. 说明如何通过限制性图谱判断连接到载体的一个限制性位点上的限制性片段的方向。请用不同于书中的片段长度。

15. 解释定点诱变的原理？叙述实施这个过程的方法。

16. 比较 S1 图谱法与引物延伸法测定 mRNA 的 5′端的异同，哪种方法还可用来测定 mRNA 的 3′端，另一种方法为什么不行?

17. 什么是截断转录，为什么这种方法不能像 S1 图谱法和引物延伸法一样用于检测体内转录物?

18. 怎样标记双链 DNA 的 5′端、3′端?

19. 描述核连缀转录分析，说明其与截断转录分析的不同。

20. 斑点杂交与 Southern 印迹法有什么不同?

21. 描述使用报告基因检测启动子的活性。

22. 描述滤膜结合实验测定 DNA 与蛋白质的结合。

23. 比较分析 DNA-蛋白质相互作用的凝胶阻滞实验和 DNase 足迹法之间的异同。DNase 足迹法提供了哪些凝胶阻滞实验不能提供的信息。

24. 比较 DMS 足迹法与 DNase 足迹法的异同。为什么前者比后者更精确?

25. 描述构造基因敲除小鼠的方法，说明胸苷激酶基因和新霉素抗性基因在这个过程中的重要性。基因敲除小鼠能提供哪些信息。

分析题

1. 对一些 DNA 片段进行了琼脂糖凝胶电泳并获得了图 5.2 所显示的结果。

a. 迁移 25mm 片段的大小是多少?

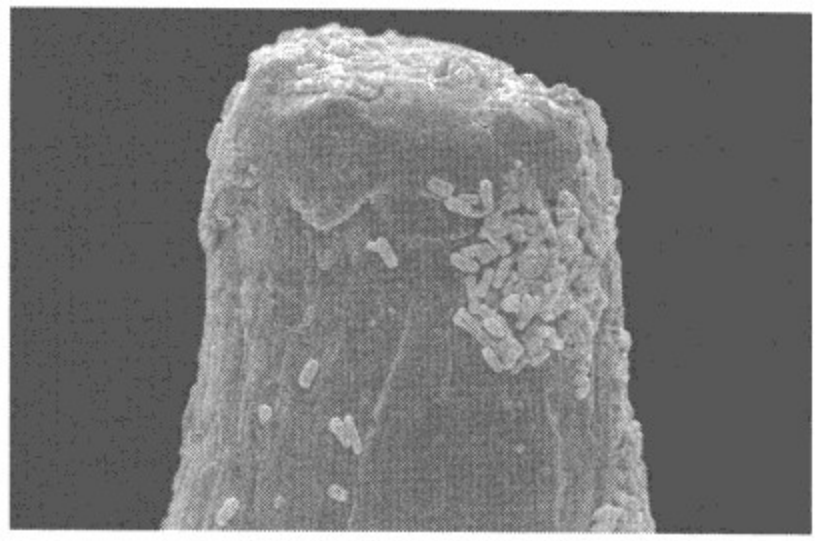
b. 200bp 的片段迁移了多远?

2. 设计一个 Southern 印迹实验，检测嵌合体小鼠的 DNA 是否插入新霉素抗性基因。可以选择靶基因和新霉素抗性基因上的任何一个限制酶位点，请显示成功插入和未成功插入抗性基因的结果。

3. 在 DNase 足迹实验中模板链或非模板链都可以被标记。在图 5.37a 中模板链是标记的，那么图 5.37b 中哪条链是标记的？你是怎么知道的?

翻译　马　纪　校译　张富春　郑用琏

第6章　细菌的转录装置



位于大头针针尖的 *E. coli* 的假彩色扫描电子显微镜图像（×1700）。
© *Andrew Syred / Stone / Getty.*

在第3章我们已经了解到转录是基因表达的第一步。确实，转录是许多基因表达的一个关键调控点。第6～9章将详细阐述细菌的基因转录及其调控机制，本章内容主要侧重于转录的基本机制，包括催化转录的 RNA 聚合酶，以及 RNA 聚合酶与 DNA 间的相互作用。当 RNA 聚合酶锚定到启动子（与基因相邻的 RNA 聚合酶特异结合位点）上时，就意味着相互作用的开始，并随着 RNA 链的延伸而持续，当 RNA 聚合酶到达终止子即终止位点时，相互作用结束，并释放合成的转录物。

6.1　RNA 聚合酶的结构

早在1960～1961年，相继在动物、植物和细菌中发现了 RNA 聚合酶。而且，正如所预期的那样，对细菌 RNA 聚合酶的研究最为详尽。到1969年，组成 *E. coli* RNA 聚合酶的多肽组分通过第5章描述的 SDS-PAGE 而确定。

图6.1是 Richard Burgess、Andrew Travers 及同事分离 *E. coli* RNA 聚合酶亚基的 SDS-PAGE 实验结果图示，该酶包含2个大亚基：β 和 β′，其分子质量分别为 150kDa 和 160kDa，在该实验中，β 和 β′亚基未能被很好地分离开来，但在随后的实验中获得了明确的分离鉴定。此外还包含 σ（sigma）亚基和 α 亚基，分子质量分别为 70kDa 和 40kDa，在 SDS-PAGE 上可观察到这两个亚基的存在。本实验未检测到分子质量为 10kDa 的 ω 亚基，但对相同

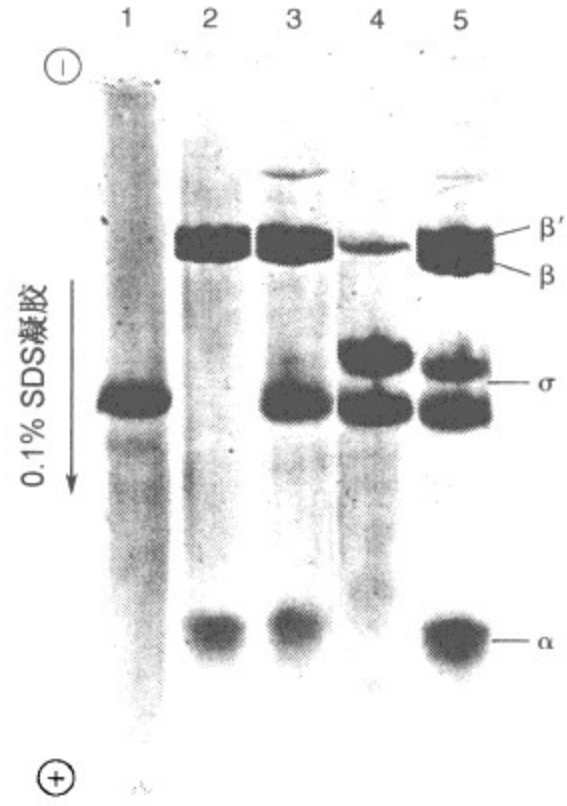


图6.1　磷酸纤维素层析法分离 *E. coli* RNA 核心聚合酶与 σ 因子。 Burgess、Travers 及同事对 *E. coli* RNA 聚合酶全酶进行磷酸纤维素层析分析，产生 A、B 和 C 3个蛋白质峰。然后，对全酶（泳道1）、峰 A、B 和 C（泳道2～4）及纯化的 σ 因子（泳道5）进行 SDS-PAGE 分析。峰 A 包含 σ 因子及一些污染物；峰 B 包含全酶；峰 C 包含功能性核心酶（亚基 α、β 和 β′）。（译者注：泳道标号从左到右应为5、4、3、2、1）（*Source*: Burgess et al., "Factor Stimulating Transcription by RNA Polymerase." *Nature* 221 (4 January 1969) p. 44, fig. 3. © Macmillan Magazines Ltd.）

的 *E. coli* RNA 聚合酶样品进行尿素-聚丙烯酰胺凝胶电泳时，检测到了该亚基的存在。与其他亚基相反，ω 亚基并不是细胞生活力所必需的，在体外也不影响酶活性。虽然不是必需亚基，但 ω 亚基可能在酶的组装过程中起作用。用星号标注的多肽（译者注：图上未给出星号）是酶的污染物。**RNA 聚合酶全酶**（RNA polymerase holoenzyme）由 β′ 亚基、β 亚基、σ 亚基、α 亚基和 ω 亚基组成，即 2 分子 α 亚基和各 1 分子其他亚基。

当 Burgess、Travers 及同事利用阴离子树脂磷酸纤维素对 RNA 聚合酶全酶进行离子交换层析（第 5 章）处理时，检测到 3 个蛋白质峰，标注为 A、B、C。用 SDS-PAGE 分析这些蛋白质峰时发现，可将 σ 亚基与全酶的其余组分，即**核心聚合酶**（core polymerase）分离开来，泳道 2 的结果表明，蛋白质峰 A 的成分含有 σ 亚基、多肽污染物以及少量的 β′ 亚基；泳道 3 为峰 B 包含的多肽，其中包括全酶；泳道 4 显示峰 C 包含的组分，包括核心聚合酶，明显缺失 σ 亚基；泳道 5 是从样品中进一步纯化 σ 亚基的结果，已去除了大多数污染物质。

接着，研究者测试了 RNA 聚合酶全酶中核心聚合酶与 σ 因子的 RNA 聚合酶活性。结果表明，两者的分离可引发酶活性的显著改变（表 6.1）。在体外，全酶对 T4 噬菌体的完整 DNA 具有很高的转录活性，而核心酶的转录活性却很低。此外，核心聚合酶仍然保持聚合 RNA 的基本功能，能够高效地转录具切口的模板（带有单链切口的 DNA）（我们将会了解到，切口 DNA 的转录是在实验室中的人为现象，无任何生物学意义）。

表 6.1　核心酶与全酶转录 DNA 的活性

DNA 模板	相对转录活性	
	核心酶	全酶
T4（天然，完整）	0.5	33.0
小牛胸腺（天然，切口）	14.2	32.8

σ 亚基是一种特异性因子

将 σ 亚基加入到核心酶又重新恢复了其转录完整 T4 DNA 的活性。更有意义的是，Ekkehard Bautz 和同事发现，全酶只转录某一类特定的 T4 基因（极早期基因），但核心酶却不具有这样的专一性。

核心酶不仅不加选择地转录 T4 噬菌体的基因，而且还转录 DNA 的两条链。Bautz 及同事用 RNA 聚合酶全酶或核心酶转录的标记产物与天然 T4 噬菌体的 RNA 分子杂交，然后检测杂交分子对 RNase 的抗性来证明了这一点（图 6.2）。即获得两条互补配对的 RNA 分子，从而形成对 RNase 有抗性的双链 RNA，由于天然 T4 RNA 是非对称（asymmetrically）转录（在任意给定的区域内，只有一条 DNA 链被转录）的产物，不应该与体外正确转录形成的 T4 RNA 杂交。因为该 RNA 也是以非对称方式转录的，与真正的 T4 RNA 相同，不能互补。Bautz 及同事以 RNA 聚合酶全酶体内转录形成的 T4 RNA 为研究对象，确实观察到了这种现象。然而，如果在体外 T4 RNA 是以对称方式合成的，则会有一半的体外转录物与体内 RNA 互补，可发生杂交而获得 RNase 抗性。事实上，Bautz 及同事发现，确实有

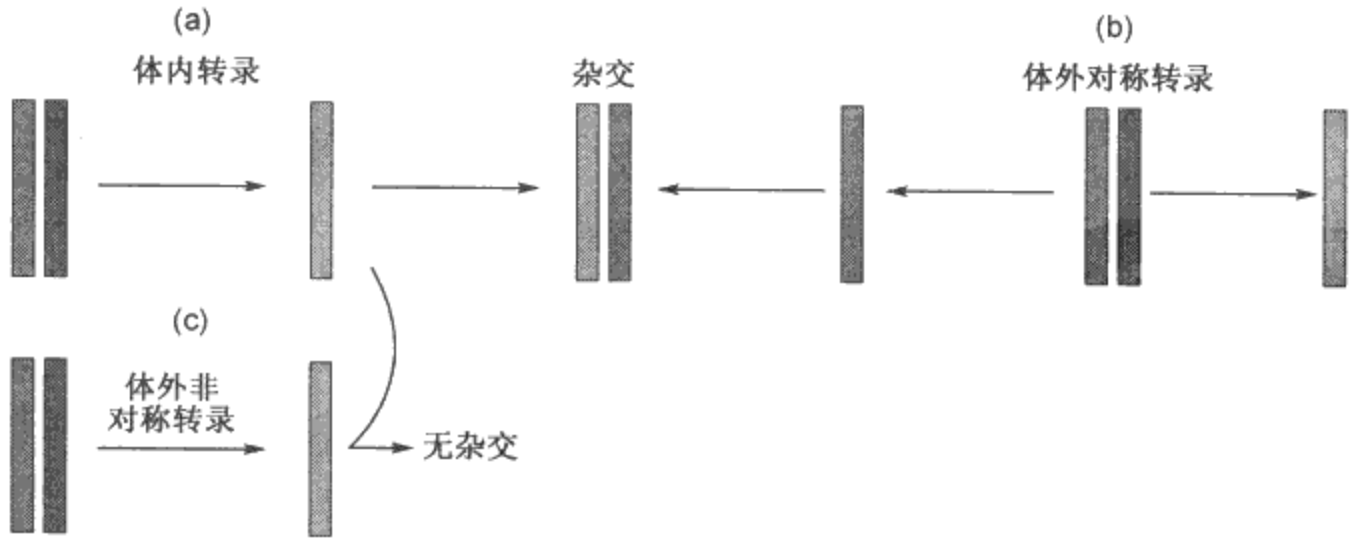


图 6.2　体外非对称性转录或对称性转录的检测。（a）在体内，转录是非对称性的，图中所示区域内只有一条 DNA 链（红色）被转录，产生 RNA（橘黄色）。（b）在体外，如果转录是对称性的，则 DNA 的两条链（红色和蓝色）均能被转录产生 RNA 转录物（分别用橘黄色和绿色表示）。产物之一（绿色）与体内的 RNA 产物（橘黄色）互补，能够发生杂交，可以利用其对 RNase 抗性而检测到杂交产物。（c）如果体外转录与体内转录的情况一样，都是非对称转录，只有一条正确的 DNA 链（红色）被转录，产生 RNA 转录物（橘黄色），这条 RNA 产物与体内的 RNA 产物是一致的，因此两者不能杂交。通过产物仍然具有对 RNase 的敏感性而证明这一点。

30%由核心聚合酶在体外转录形成的标记 RNA 与天然 T4 RNA 杂交后，获得 RNase 抗性，这说明，在体外核心酶以非正常的方式转录 DNA 双链。

很显然，全酶失去 σ 亚基后，核心酶只具有基本的 RNA 合成能力，不具有特异性。重新结合 σ 亚基后又恢复其特异性。实际上，在明了这一特性后才将该因子命名为 σ 因子，选用 σ 或希腊字母 s 来表示特异性。

小结 RNA 聚合酶是转录过程中的关键因子。*E. coli* 的 RNA 聚合酶是由核心酶和 σ 亚基组成，核心酶是基本的转录装置，σ 因子能够指导核心酶转录特异的基因。

6.2 启动子

在表 6.1 提及的 T4 DNA 转录实验中，为什么核心聚合酶具有转录切口 DNA 的能力，而不能转录完整的 DNA？这是因为 DNA 分子上的切口（nick）或间隙（gap）为 RNA 聚合酶，甚至核心酶提供了理想的转录起始位点。但这种转录起始必定是非特异性的。在完整 T4 DNA 中几乎没有切口或间隙，核心聚合酶也就很难遇到这种人为的起始位点，因此对其只有微弱的转录。此外，如果 σ 亚基存在，RNA 聚合酶全酶就能识别天然 T4 DNA 上真正的聚合酶结合位点，并在该位点起始转录。聚合酶结合的位点称为启动子（promoter）。在启动子处起始的体外转录是特异的，如同在体内发生的转录。所以，σ 因子的作用就是指导聚合酶在特异的启动子处起始 DNA 的转录。在本节我们将学习细菌聚合酶与启动子间的相互作用以及这些启动子的结构。

RNA 聚合酶与启动子的结合

σ 因子如何改变核心酶对启动子结合的行为方式？David Hinkle 和 Michael Chamberlin 用硝化纤维素滤膜结合实验（第 5 章）回答了这个问题。为检测全酶及核心酶与 DNA 结合的紧密程度，他们从 *E. coli* 中分离出这些酶，并与 ^{3}H 标记的 T7 噬菌体 DNA 结合，该 DNA 的早期启动子可被 *E. coli* 聚合酶所识别。加入过量未标记的 T7 DNA，以便从标记的 T7 DNA 解离的聚合酶与未标记 DNA 结合的概率远高于标记的 DNA。在反应不同时间后，将混合物用硝化纤维素滤膜处理。只有仍与聚合酶结合的标记 DNA 可结合在膜上。因此，该实验能够测定聚合酶-DNA 复合体的解离速率。当最后一个聚合酶（可能是结合最紧的）从标记的 DNA 上脱解离下来，标记的 DNA 也就不再与滤膜结合，滤膜的放射信号减弱。

图 6.3 显示了实验结果。显然，RNA 聚合酶全酶较核心酶对 T7 DNA 的结合更紧密。事实上，全酶解离的半衰期（$t_{1/2}$）为 30～60h，即经过 30～60h 后，只有一半的聚合酶-DNA 复合物解离，表明聚合酶全酶与 DNA 的结合确实十分紧密。相比之下，核心酶解离的 $t_{1/2}$ 小于 1min，相对于聚合酶全酶而言，核心酶与 DNA 的结合要松散得多。因此，σ 因子至少是在某些特定 DNA 位点处促进聚合酶与 DNA 的紧密结合。

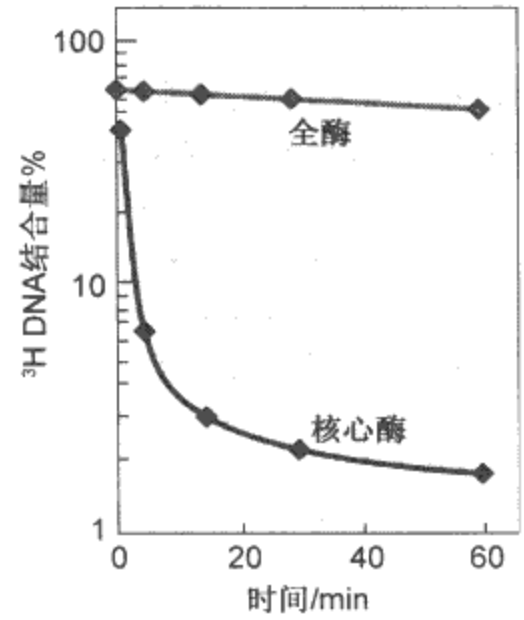


图 6.3 σ 因子促进 RNA 聚合酶与启动子的紧密结合。Hinkle 和 Chamberlin 使 ^{3}H 标记的 T7 DNA 与 *E. coli* RNA 聚合酶核心酶（蓝色）或全酶（红色）结合，然后加入过量的未标记 T7 DNA，这样从标记 DNA 上脱离下来的聚合酶就可能与未标记的 T7 DNA 重新结合。将反应了不同时间的混合物通过硝化纤维素滤膜，检测标记 T7 DNA-聚合酶复合物的解离（随着最后一个聚合酶与标记 DNA 的解离，DNA 不再与滤膜结合，滤膜的放射性随之消失）。相对于核心聚合酶（蓝色）而言，全酶（红色）与 T7 DNA 的解离速率要慢得多，表明全酶与 T7 DNA 结合得非常紧密。（*Source*: Adapted from Hinckle, D. C. and Chamberlin, M. J., "Studies of the Binding of *Escherichia coli* RNA Polymerase to DNA," *Journal of Molecular Biology*, Vol. 70, 157-85, 1972.）

在解离实验中，Hinkle 和 Chamberlin 转换了实验程序，先使聚合酶与未标记 DNA 结合，然后加入过量的标记 DNA，最后，在不同时间内将混合物通过硝化纤维素滤膜。这一过程测定了最初（和最松散结合）的聚合酶解离情况，因为新解离的聚合酶能够与游离的标记 DNA 结合，从而使标记 DNA 结合到滤膜上。此分析揭示，全酶与核心酶一样，

在 DNA 上都有松散结合位点。

Chamberlin 解释认为，全酶在 T7 DNA 上存在两种结合位点：紧密结合位点和松散结合位点。而核心聚合酶只能与 DNA 松散结合。因为 Bautz 等已经证明是全酶而不是核心酶能够识别启动子，由此推测紧密结合位点可能就是启动子，而松散结合位点则是 DNA 的其他部分。Chamberlin 及同事的研究也表明，全酶与 T7 DNA 的紧密结合复合物在添加核苷酸时能立即起始转录，强化了紧密结合位点确实是启动子的结论。如果聚合酶已经与远离启动子的位点发生了紧密结合，那么，会因为聚合酶寻找起始位点而发生转录滞后现象。进而，Chamberlin 及同事用滴定法分析聚合酶在 T7 DNA 分子上的紧密结合位点，只发现 8 个位点，与该 DNA 的早期启动子数目相近。相反，全酶与核心酶在 DNA 上的松散结合位点大约有 1300 个，这些松散结合位点出现在 DNA 的任何区域，因此也就没有特异性。核心酶不能与紧密结合位点（启动子）结合，这也解释了核心酶不能特异地起始 DNA 转录的原因，因为起始转录需要与启动子的紧密结合。

Hinkle 和 Chamberlin 也检验了温度对全酶与 T7 DNA 结合的影响，发现提高温度能显著增强两者间的紧密结合。图 6.4 显示，在 25℃时的解离速率明显高于 37℃时的解离速率，15℃时解离速率更高。因为高温能促进 DNA 熔解（解链，第 2 章），这一发现与紧密结合涉及 DNA 局部解链的观点一致。在本章的后面部分我们将介绍这一假设的直接证据。

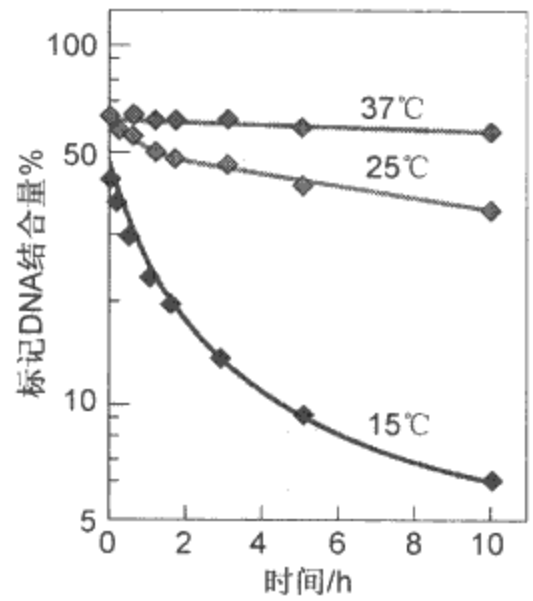


图 6.4　温度对全酶-T7 DNA 复合体解离的影响。Hinkle 和 Chamberlin 使 *E. coli* RNA 聚合酶全酶与^{3}H 标记的 T7 DNA 在 37℃（红色）、25℃（绿色）、15℃（蓝色）条件下结合形成复合体。然后，加入未标记的 T7 DNA，使之与解离的聚合酶发生竞争性结合；在不同时间取样并进行硝化纤维素滤膜处理以监测聚合酶-标记 T7 DNA 复合物的解离情况。37℃条件下形成的复合体比 25℃条件下形成的复合体稳定得多；而后者又比 15℃条件下形成的复合体稳定。说明较高温度有利于 RNA 聚合酶全酶与 T7 DNA 的紧密结合。（*Source*: Adapted from Hinckle, D. C. and Chamberlin, M. J., "Studies of the Binding of *Escherichia coli* RNA Polymerase to DNA," *Journal of Molecular Biology*, Vol. 70, 157-85, 1972.）

Hinkle 和 Chamberlin 将有关聚合酶与 DNA 之间相互作用的研究发现总结为以下假说（图 6.5）：RNA 聚合酶全酶先与 DNA 松散结合，可能一开始就结合在启动子上，或沿 DNA 链扫描直到发现启动子。全酶与启动子松散结合形成的复合物称为闭合启动子复合体（closed promoter complex），因为 DNA 依然保持着闭合的双链形式。然后，全酶使处于启动子内的一小段区域熔解而形成**开放的启动子复合体**（open promoter complex），这时聚合酶与 DNA 间紧密结合。之所以称为开放的启动子复合体，是因为该开放复合体的形成依赖于 DNA 的解旋。

由聚合酶在闭合启动子复合体中的松散结合转变为在开放启动子复合体中的紧密结合需要 σ 因子的参与，并导致转录的开始。现在，我们可以领悟 σ 因子在决定转录特异性中是如何发挥功能的：σ 因子选择能与 RNA 聚合酶紧密结合的启动子，这样与启动子相邻的基因将被转录。

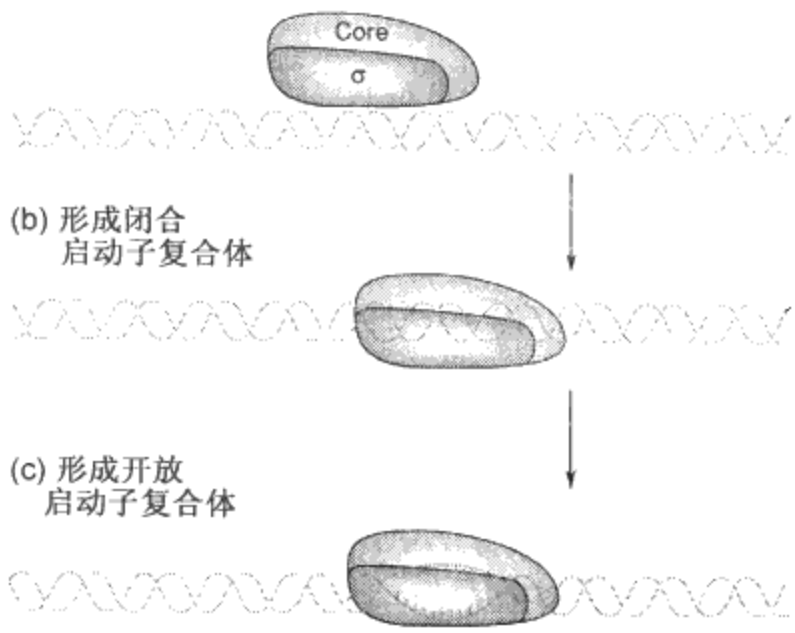


图 6.5　RNA 聚合酶与启动子结合。（a）RNA 聚合酶全酶与 DNA 进行松散地结合与再结合，并寻找启动子；（b）全酶找到启动子并与之松散结合，形成闭合启动子复合体；（c）全酶的紧密结合使 DNA 局部解链，形成开放的启动子复合体。

小结　σ 因子通过引起 RNA 全酶紧密地结合到启动子上而引发转录的起始。这种紧密结合

依赖于 DNA 局部解链形成开放启动子复合体，这一事件由 σ 因子促进。因此，σ 因子能够选择将被转录的基因。

启动子结构

细菌启动子具有怎样的特性使 RNA 聚合酶与之结合？David Pribnow 对 *E. coli* 和噬菌体的多个启动子加以比较后，辨别出一个共有区域：其中心位于转录起始点上游约 10bp 处一段 6～7bp 的序列，最初被称为“Pribnow 框”（Pribnow box），现在常称为－10 框（－10 box）。Mark Ptashne 及同事注意到还存在另一个短序列：其中心位于转录起始位点上游约 35bp 处，被称为－35 框（－35 box）。分析数千个启动子后发现，每个盒都存在典型或共有序列（consensus sequence）（图 6.6）。

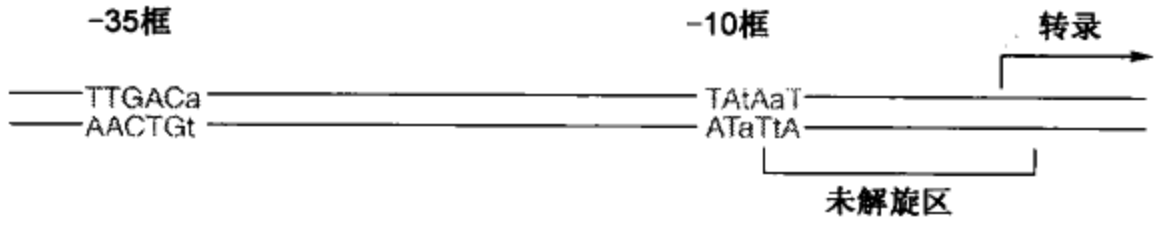


图 6.6　细菌启动子。在一个典型 *E. coli* 启动子中，分别标注了相对于转录起始点的－10 框、－35框以及未解旋区域的位置。大写字母表示该位置碱基在已研究的启动子中出现的频率大于 50%；小写字母表示该位置碱基在已研究的启动子中出现的频率等于或小于 50%。

这些所谓的共有序列仅代表一种概率。图 6.6 中大写字母碱基表示在该位置出现的频率较高，小写字母碱基表示在该位置也经常出现，但频率比大写字母碱基要低。这种概率表示很难找到与共有序列完全匹配的－10 框与－35 框。然而，完全匹配序列往往出现在极强启动子中，这些启动子异常活跃地启动转录。事实上，破坏与共有序列匹配率的突变趋向于**下降突变**（down mutation），即突变使启动子的活性减弱而导致转录减少。使启动子序列更接近于共有序列的突变则会使启动子的功能增强，这种突变称为**上升突变**（up mutation）。启动子元件之间的距离也十分重要，使－10 框与－35 框非正常地靠近或远离的删除或插入突变都是有害的。在第 10 章中我们将了解到真核生物的启动子也具有其共有序列，其中之一与－10 框十分类似。

除了称之为**核心启动子元件**（core promoter element）的－10 框和－35 框外，有些极强启动子在更远的上游还存在一个额外元件，称为 **UP 元件**（UP element）。*E. coli* 细胞有 7 个编码 rRNA 的基因（*rnn* gene），当快速生长需要大量 rRNA 时，这 7 个基因在细胞内自身就可完成大量的转录。很显然，驱动这些基因转录的启动子非常强大，启动子的上游元件可以部分解释这一事件。图 6.7 显示了其中一个启动子 *rrnB* P1 的结构。核心启动子（蓝色）上游－40～－60 处有一个 UP 元件（红色）。我们之所以认为 UP 元件是一个真正的启动子元件，是因为仅在 RNA 聚合酶存在的情况下它能将 *rrnB* P1 基因的转录效率提高 30 倍，加之可被 RNA 聚合酶自身所识别，所以，我们断定 UP 元件是一个启动子元件。

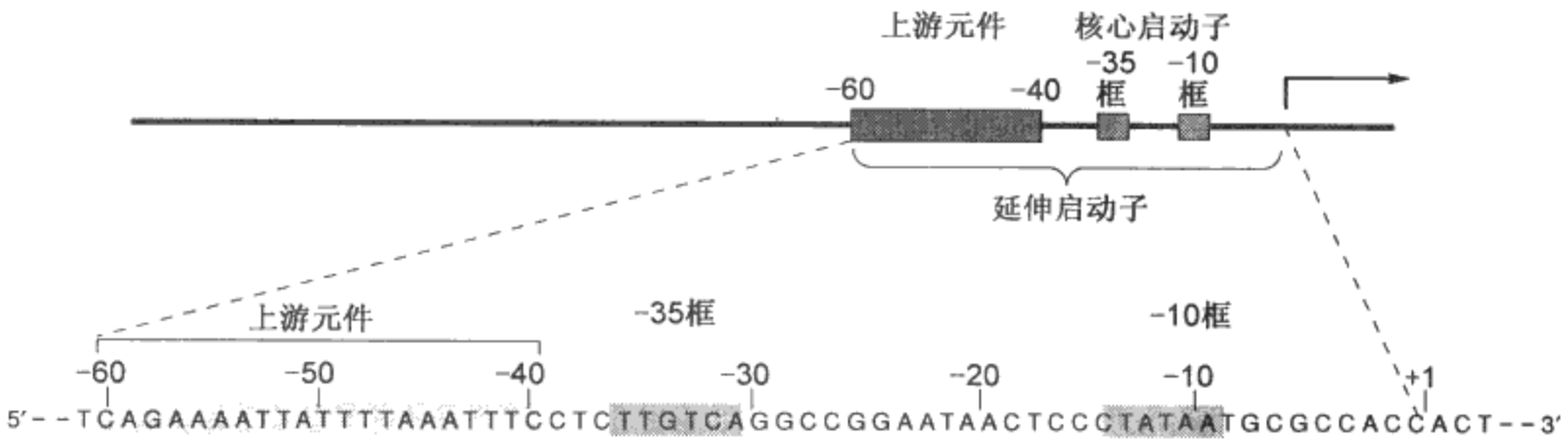


图 6.7　*rrnB* P1 启动子。核心启动子（－10 框与－35 框，蓝色）和上游是 UP 元件（红色）在图上方标出；示意图下方以相同的颜色标出了这些元件的全部碱基序列（非模板链）。（*Source*：Adapted from Ross et al.，“A third recognition element in bacterial promoters：DNA binding by the alpha subunit of RNA polymerase.” *Science* 262：1407，1993.）

该启动子还与－60 与－150 间的 3 个 *Fis* 位点有关，是转录激活蛋白 Fis 的结合位点。这些位点本身不与 RNA 聚合酶结合，所以不是典型的启动子元件，而是一类被称为**增强子**（enhancer）的转录激活 DNA 元件。我们将在第 9 章详细讨论细菌增强子。

E. coli rrn 启动子也能够被一对小分子物质所调节，即**起始 NTP**（initiating NTP，**iNTP**）和**预警子**（alarmone）鸟苷 5′-二磷酸-3′-二磷酸（ppGpp）。大量 iNTP 的存在表明核苷酸的浓度很高，有利于合成大量的 rRNA。相应地，iNTP 能够稳定开放启动子复合体而促进转录。

此外，当细胞发生氨基酸饥饿时，蛋白质的合成不会轻易发生，对核糖体（和 rRNA）的需要量也随之减少。核糖体位点是氨酰-tRNA 的正常结合位点，当空载 tRNA 与核糖体位点结合后，核糖体对氨基酸缺乏十分敏感。在这些情况下，核糖体相关蛋白 RelA 接收“预警”，催化合成“预警子”ppGpp。预警子可破坏寿命较短的开放启动子复合体的稳定性，从而抑制转录。

蛋白质 DskA 与 RNA 聚合酶结合，将 *rrn* 开放启动子的寿命降低到能够对 iNTP 和 ppGpp 浓度的改变做出反应的水平。因此，DskA 蛋白是 iNTP 和 ppGpp 调节 *rrn* 转录所必需的。事实上，在 DskA 缺失突变体中，*rrn* 的转录确实对 iNTP 和 ppGpp 不敏感。

小结 细菌启动子包含两个区域，其中心分别位于转录起始点上游的－10bp 和－35bp 处。在*E. coli*中，这两个区域或多或少分别与 2 个保守序列 TATAAT 和 TTGACA 相似。一般来讲，启动子内的这些区域与保守序列越相似，启动子起始转录的能力就越强。某些极强启动子在核心启动子上游还包含一个额外元件（UP 元件），使启动子对 RNA 聚合酶具有更强的吸引力。由 *rrn* 启动子处起始的转录受 iNTP 浓度的正调控，受 ppGpp 的负调控。

6.3 转录起始

在 1980 年之前，普遍认为当 RNA 聚合酶催化形成第一个磷酸二酯键，最初的 2 个核苷酸加入到生长的 RNA 链之后，转录的起始就结束了。Agamemnon Carpousis 和 Jay Gralla 研究指出，转录起始实际上比这更复杂。研究者将 *E. coli* RNA 聚合酶与含 *E. coli lac* UV5 启动子的 DNA 一起孵育，同时加入肝素。肝素是一种带负电荷的多聚糖，可与 DNA 竞争结合游离的 RNA 聚合酶。转录循环结束后，肝素能够阻止释放出的 RNA 聚合酶与 DNA 重新结合。用标记的 ATP 标记 RNA 产物，然后电泳产物，测定其大小，发现存在一些很小的寡核苷酸，其大小从二聚体到六聚体（2～6nt），如图 6.8 所示。这些寡核苷酸序列与预期的 *lacZ* 启动子转录的起始序列相一致。而且，当 Carpousis 和 Gralla 测定这些寡核苷酸的数量并与 RNA 聚合酶的数量比较时发现，每个 RNA 聚合酶对应于多个寡核苷

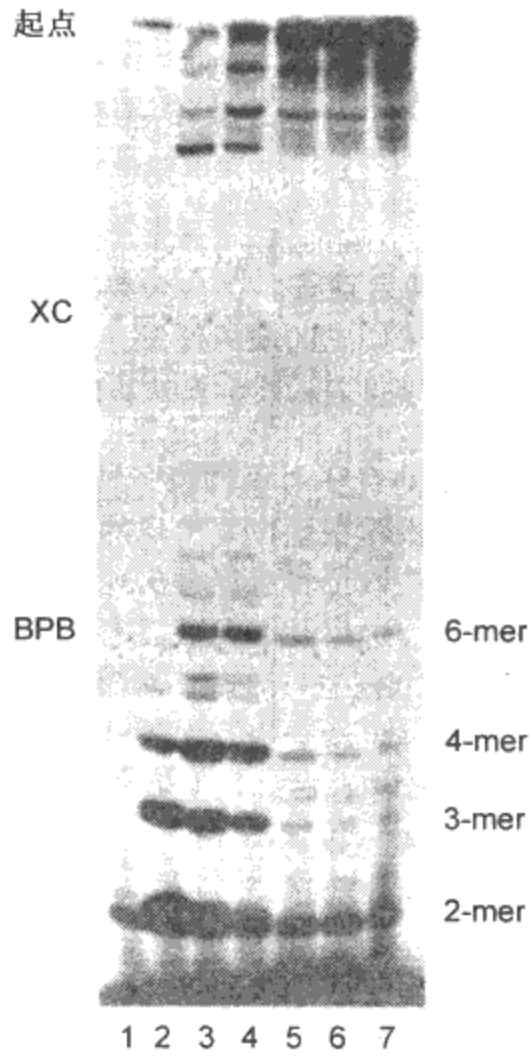


图 6.8 RNA 聚合酶结合在启动子上合成短的寡核苷酸片段。Carpousis 和 Gralla 在体外用 RNA 聚合酶合成^{32}P 标记的 RNA，反应体系包含携带 *lac* UV5 启动子的 DNA 模板、可结合游离 RNA 聚合酶的肝素、[^{32}P] ATP、不同浓度的其他 3 种脱氧核苷酸（CTP、GTP 和 UTP）。通过聚丙烯酰胺凝胶电泳和放射自显影检测标记的 RNA 产物。泳道 1：不含 DNA 的对照；泳道 2：只有 ATP；泳道 3～7：ATP 以及浓度不断加倍的 CTP、GTP 和 UTP，即从泳道 3 的 25μmol/L 到泳道 7 的 400μmol/L。二聚体到六聚体的位置在图右侧标出。两种标记染料溴酚蓝（BPB）和二甲苯（XC）的位置在图左侧标出。泳道 1 中出现的二聚体可能是标记 ATP 中的污染物，该污染物同样出现在泳道2～7 中。（*Source*: Carpousis A. J. and Gralla J. D. Cycling of ribonucleic acid polymerase to produce oligonucleotides during initiation in vitro at the *lac* UV5 promoter. *Biochemistry* 19（8 Jul 1980）p. 3249, f. 2 © American Chemical Society.）

酸。由于肝素阻止了游离 RNA 聚合酶与 DNA 的再结合，因此该结果暗示 RNA 聚合酶甚至在没有离开启动子时，就产生了很多小的**流产性转录物**（abortive transcript）。其他研究者也证实了这一结果，并发现了长度达 9nt 或 10nt 的流产性转录物。

至此，我们已经了解到转录起始过程比最初设想的要复杂得多。如图 6.9 所示，现在一般认为转录起始可分为 4 个步骤：①闭合启动子复合体形成；②闭合启动子复合体向开放启动子复合体转换；③聚合最初几个核苷酸（可达 10 个）时，RNA 聚合酶仍停留在启动子处；④启动子清除（promoter clearance）。此时转录物达到足够长度并与 DNA 模板链形成稳定的杂合体，有助于稳定转录复合体，RNA 聚合酶转变成延伸构象，释放 σ 因子，移出启动子区。本节将详尽阐述转录起始过程。

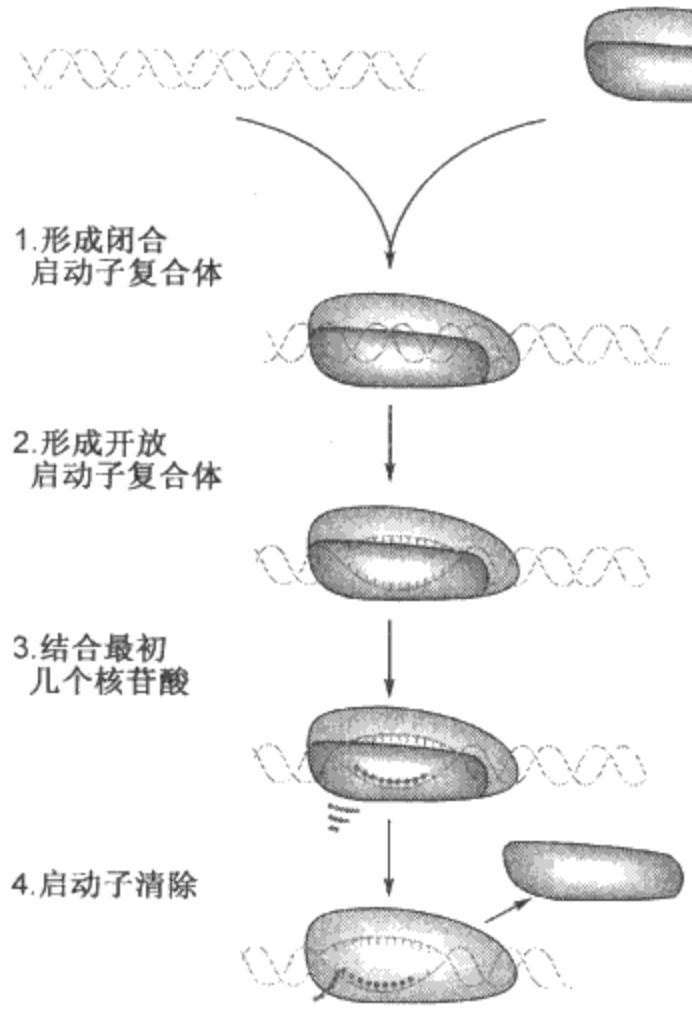


图 6.9　转录起始的阶段。1. RNA 聚合酶结合在 DNA 上形成闭合启动子复合体。2. σ 因子使聚合酶由闭合启动子复合体转换为开放启动子复合体。3. 聚合酶合成 9～10nt 的初生 RNA 产物。图左侧为流产性转录物。4. 聚合酶的启动子清除，释放 σ 因子，进入转录延伸阶段。

σ 因子的功能

我们已经了解到，σ 因子通过引起 RNA 聚合酶与启动子间的紧密结合来选择被转录的基因，而这种紧密结合依赖于 DNA 的局部解链，进而形成开放启动子复合体。我们也知道，在完成聚合酶与启动子的结合后，σ 因子与核心酶解离，这意味着 σ 因子能促进转录起始，并不断被重复利用。现在，我们将给出这些结论的证据。

σ 因子促进转录起始　因为 σ 因子指导 RNA 聚合酶与启动子的紧密结合，将酶置于启动转录的位置，恰好位于基因的起始端。所以，我们预想 σ 因子能够激活转录起始。为验证这一点，Travers 和 Burgess 利用 RNA 合成时加入的第一个核苷酸保留其携带的 3 个磷酸基团（α、β 和 γ），而 RNA 链中其他核苷酸只保留其 α-磷酸（第 3 章）特点，在逐渐增加 σ 因子浓度的条件下，与 RNA 聚合酶的核心酶共同孵育，进行了两组独立实验。在一组反应中，被标记核苷酸是［^{14}C］ATP，在整条 RNA 链中均有掺入，因此可用于检测 RNA 链转录的起始和延伸；在另一组反应中，被标记的核苷酸是［γ-^{32}P］ATP 或［γ-^{32}P］GTP，该标记核苷酸只出现在 RNA 链的第一个位置处，用于检测转录的起始（研究者之所以使用 ATP 或 GTP，是因为转录起始的第一个碱基通常是嘌呤核苷酸，且 ATP 较 GTP 更为常见）。如图 6.10 所示，研究结果表明，σ 因子同时促进了^{14}C 和 γ-^{32}P 标记核苷酸的掺入，表明 σ 因子既能促进转录起始，又能促进转录延伸。然而，起始反应是转录的限速步骤（起始新 RNA 链的合成比延伸该 RNA 链要花费更长时间）。因此，σ 因子似乎是通过促进转录起始从而促进链的延伸，为核心酶提供了更多用于延伸的起始链。

Travers 和 Burgess 通过证明 σ 因子并没有真正加速 RNA 链的生长速率而证实了这一点。保持 RNA 链的数目不变，证明在此条件下 σ 因子并不会影响 RNA 链的长度。其做法是，先使一定量的转录起始事件发生而获得恒定数目的 RNA 链，然后用抗生素利福平（rifampicin）阻断其他任何链的起始，利福平能阻断细菌的转录起始，但不能阻断转录延伸。然后用超速离心检测 σ 因子有或无时所转录的 RNA 长度。结果表明，不管 σ 因子存在与否，RNA 链的长度没有差异。如果 σ 因子确实促进 RNA 链的延伸，那么存在 σ 因子时，RNA 链将会变得更长。因此，σ 因子不能促进转录的延伸。在以往的实验中，σ 因子促进延伸的表象只是促进转录起始的间接效应而已。

小结　σ 因子能够促进转录的起始，而不是延伸。

σ 因子的重复利用　在 1969 年发表的同一篇论文中，Travers 和 Burgess 证明 σ 因子可被再循环利用。该实验的关键是在低离子强度下进行转录反应，因为低离子强度能防止 RNA 核心聚合酶在基因的终点处与 DNA 模板解离，致使转录起始（通过检

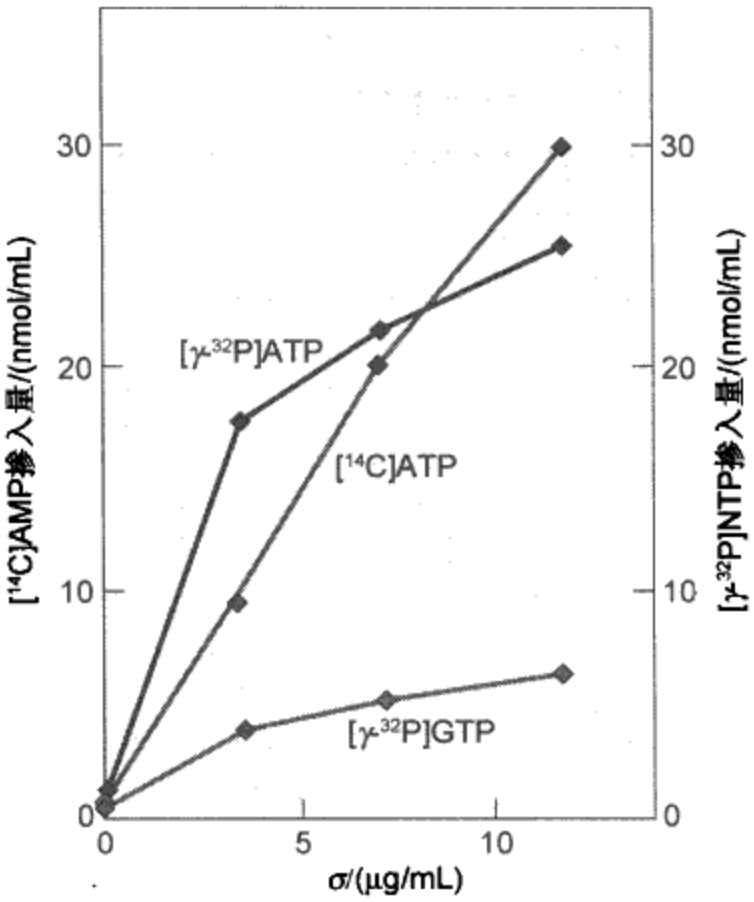


图 6.10　σ 因子几乎既能促进转录起始，又能促进转录延伸。Travers 和 Burgess 在向反应混合体系中分别添加［^{14}C］ATP（红色）、［γ-^{32}P］ATP（蓝色）或［γ-^{32}P］GTP（绿色）的条件下，通过逐渐增加 σ 因子的量，使 *E. coli* RNA 核心聚合酶在体外转录 T4 DNA。［^{14}C］ATP 的掺入量表明 RNA 在大量合成或延伸；γ-^{32}P 标记核苷酸的量可测定转录起始。由于所有浓度曲线都是上升的，所以该实验结果表明，σ 因子既能促进转录起始，又能促进转录延伸。（*Source*：Adapted from Travers，A. A. and R. R. Burgess，"Cyclic re-use of the RNA polymerase sigma factor." *Nature* 222：537-40，1969.）

测 RNA 中 γ-^{32}P 标记嘌呤核苷酸掺入的量而测定转录的起始事件）慢慢停止，如图 6.11 所示（红色曲线）。当加入新的核心酶时，发现转录重新开始（蓝色曲线），这意味着新添加的核心酶与脱离原酶的 σ 因子结合。在另一组实验中，研究者证明加入不同的 DNA 模板及核心酶后，均有新转录事件的发生，这一结果支持了以下结论：σ 因子已经从原来的核心酶中释放出来，并与新加入的核心酶在新的 DNA 模板上结合。因此，Travers 和 Burgess 提出 σ 因子可从一个核心酶到另一个核心酶进行循环利用，如图 6.12 所示，研究者称之为"σ 因子循环"。

图 6.11 还包含另一条有价值的信息。当 Travers 和 Burgess 向反应体系中加入利福平以及来自利福平抗性突变体的核心聚合酶时，转录仍能发生（绿色曲线）。由于 σ 因子来自对利福平敏感的聚合酶，所以新转录事件的利福平抗性一定是新加入

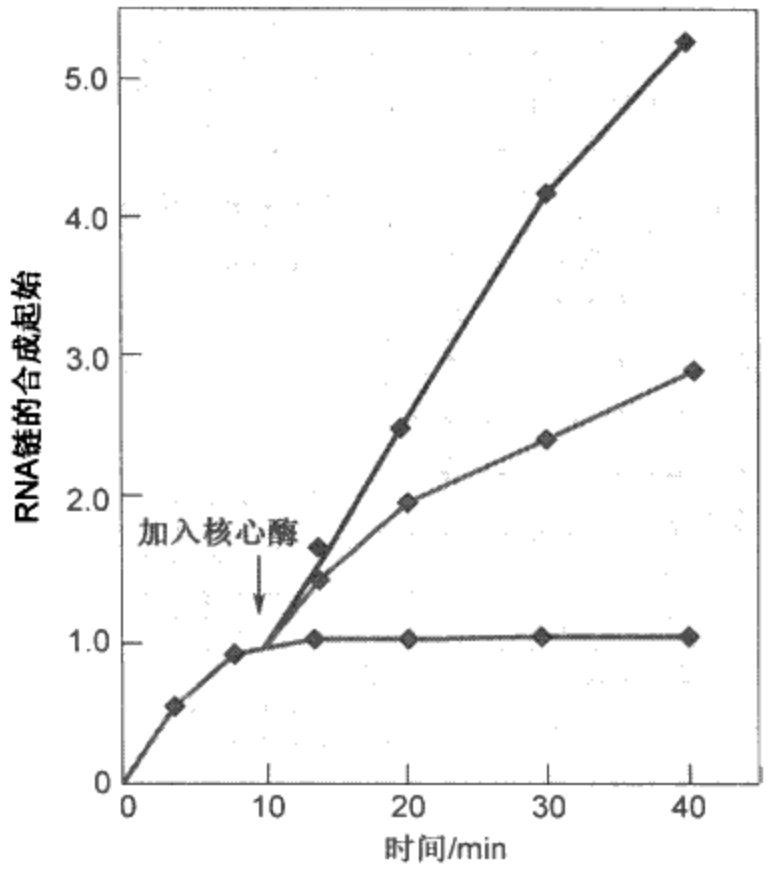


图 6.11　σ 因子能够被循环利用。在低离子强度下，Travers 和 Burgess 以 T4 DNA 为模板，使 RNA 聚合酶全酶起始转录并进入延伸阶段，当转录终止后，聚合酶不会与模板上脱离，也就无新转录事件的发生。通过测定［γ-^{32}P］ATP 和［γ-^{32}P］GTP 的掺入量可以测定 RNA 链的转录起始，如红色曲线所示。10min（箭头）后，几乎不再有转录事件的发生，此时在利福平存在（绿色曲线）或不存在（蓝色曲线）的两种情况下，向反应体系中添加新的对利福平有抗性的核心酶，结果显示，RNA 的合成又可重新开始，这说明从核心酶脱离的 σ 因子能够与新的核心酶组装成全酶，σ 因子被循环利用。同时也说明是核心酶而不是 σ 因子决定了 RNA 聚合酶全酶对利福平的抗性或敏感性。（*Source*：Adapted from Travers，A. A. and R. R. Burgess，"Cycle re-use of the RNA polymerase sigma factor." *Nature* 222：537-40，1969.）

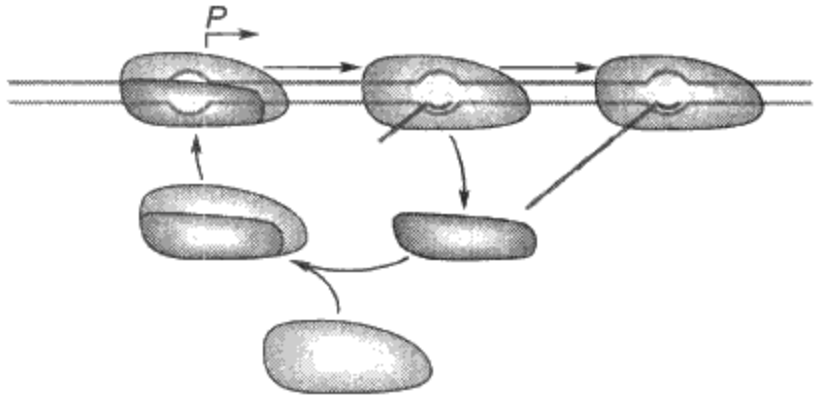


图 6.12　σ 因子循环。RNA 聚合酶与左侧的启动子结合，使 DNA 发生局部熔解。当聚合酶向右移动延伸 RNA 链时，σ 因子与核心酶分离，与新的核心酶（左下方）结合，起始另一条 RNA 的合成。

的核心酶所赋予的。事实是，在利福平存在的情况下，转录起始事件的发生有所减少，这可能意味着对抗利福平的核心酶仍然存在一定程度的敏感性。我们本以为是 σ 因子而非核心酶决定着 RNA 聚合酶对利福平的抗性或敏感性，因为利福平能够阻断转录起始，而 σ 因子又是公认的起始因子。尽管如此，核心酶才是决定利福平敏感性的关键，本章后面介绍的有关实验将对此做进一步解释。

小结 σ 因子在参与完成转录起始后，显然与核心聚合酶发生了解离，由核心酶单独执行延伸功能。而 σ 因子可被不同的核心聚合酶循环利用。而且，是核心酶而不是 σ 因子决定了 RNA 聚合酶全酶对利福平的抗性或敏感性。

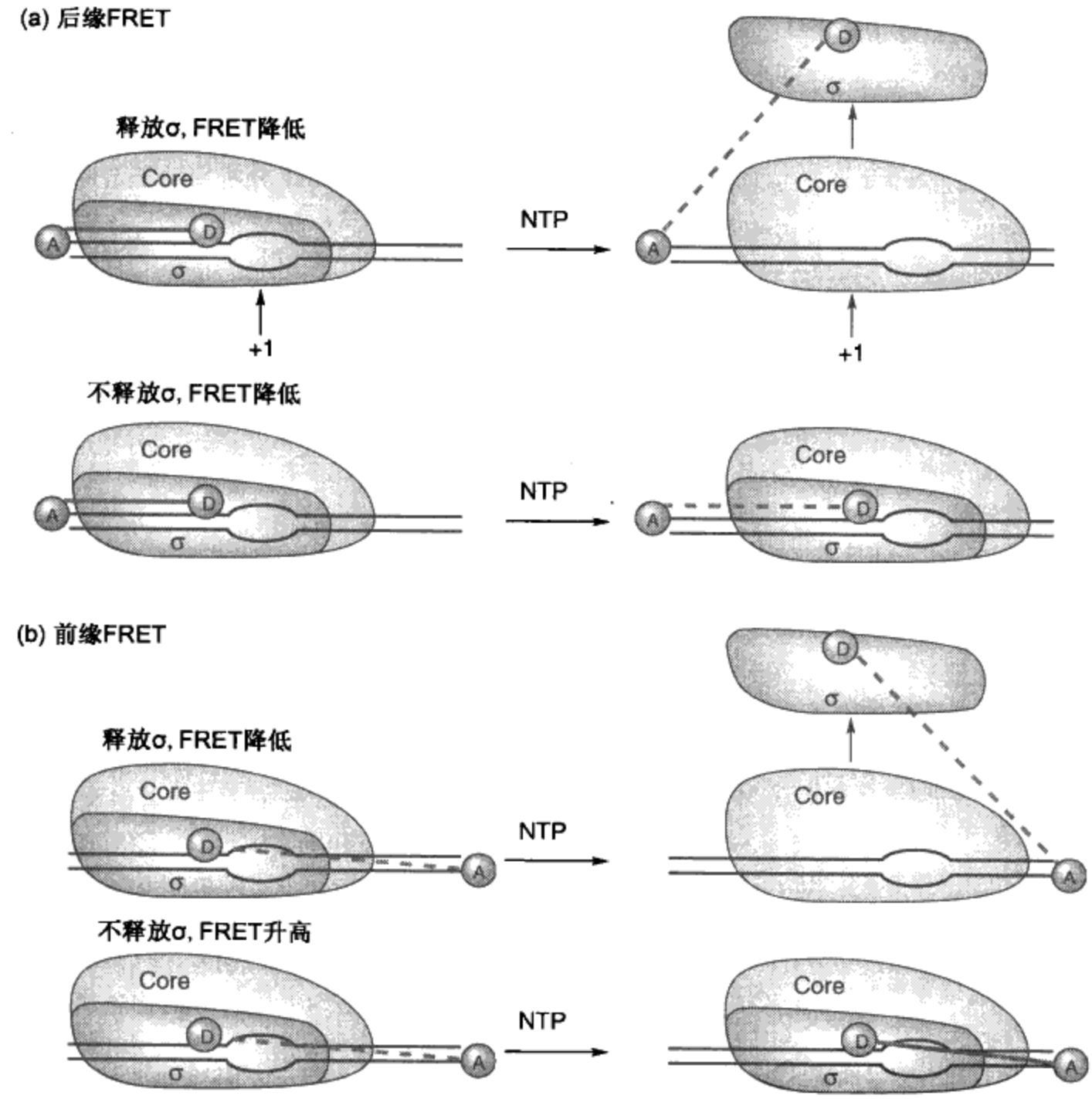


图 6.13 FRET 分析测定 σ 因子相对于 DNA 移动的基本原理。（a）后缘 FRET。荧光供体（D，绿色）附着在 σ^{70} 突变体的唯一半胱氨酸残基上，该 σ^{70} 突变体只含有一个半胱氨酸而去除了其他所有半胱氨酸。荧光受体（A，红色）附着在 DNA 的 5′端。由于两个荧光探针距离较近，在开放启动子复合体（RP_o）中 FRET 效率很高（紫色实线）。加入 4 种核苷酸中的 3 种之后，RNA 聚合酶移动到下游的某个位置，此处需要第 4 种核苷酸（CTP）。该位置至少在+11 处，所以发生启动子清除，此时，不管 σ 因子是否从核心酶上脱离，FRET 效率均降低（紫色虚线），因为两个探针相互远离。如果 σ 不解离，在延伸时它会与核心酶一起向下游移动，与 DNA 5′端的探针远离。如果 σ 解离，则 σ 因子会出现在溶液的任意位置，但一般来说，相对于转录开始前在 σ 因子开放启动子复合体中与核心酶的距离而言，σ 因子与核心酶相距更远。（b）前缘 FRET。荧光供体仍附着在 σ^{70} 的半胱氨酸残基上，但荧光受体附着在 DNA 的 3′端。由于两个探针距离较远，在开放启动子复合体中 FRET 效率很低（紫色虚线）。随着核苷酸的加入，聚合酶经过启动子清除向下游延伸到图（a）所示的位置。现在可以根据 FRET 区分这两种假说了，如果 σ 因子与核心酶解离，FRET 效率就会降低，如同在图（a）中一样；如果 σ 因子不与核心酶解离，随着 RNA 聚合酶的移动，两个探针的距离越来越近，FRET 效率随之升高（紫色实线）。

σ因子或许在延伸阶段并不与核心酶分离 经典的σ因子模型指出，聚合酶在完成启动子清除、由起始模式转换为延伸模式时，σ因子与核心酶解离。尽管这一模型提出了三十多年，并获得了大量实验证据的支持，但并非与所有的实验结果都吻合。例如，1996 年 Jeffrey Roberts 及同事研究表明，σ因子参与了在 λ 噬菌体晚期启动子 P'_R 下游＋16/＋17处的暂停，暗示 σ 因子在＋16/＋17 处仍结合在核心酶上，而此时启动子清除早已发生。

2001 年，Richard Ebright 及同事注意到所有支持 σ 因子循环模型的证据都是通过较为粗糙的分离技术获得的，如电泳或层析。在延伸阶段，如果 σ 因子与核心酶松散结合，这些分离方法就会造成 σ 因子与核心酶的解离。这样，从表面看来，在启动子清除时 σ 因子已与核心酶解离。Ebright 及同事还注意到以前的研究一般都不能区分有活性和无活性的 RNA 聚合酶。这是真正值得关注的焦点，因为在任何一个群体内，RNA 聚合酶分子内的重要片段都不具有完成由起始向延伸转换的能力。

为了验证 σ 因子循环假说，Ebright 及同事采用了**荧光共振能量转移**（fluorescence resonance energy transfer，FRET）技术，该技术无须分离就可测定 σ 因子相对于 DNA 某一位点的距离，而分离技术本身就可能引起 σ 因子与核心酶的脱离。FRET 技术的原理是两个荧光分子彼此靠近时会引发共振能量转移，这种共振能量转移的效率会随着两个荧光分子的彼此远离而迅速降低。

Ebright 及同事测定了分别结合在 σ 因子和 DNA 上的两个荧光分子（荧光探针）间的 FRET。结合在 σ 因子上的探针作为荧光供体，结合在 DNA 上的探针作为荧光受体，有时结合在 DNA 上荧光探针位于 5′端或上游末端（后缘 FRET），便于研究者在聚合酶远离启动子和 DNA 5′端时测定 FRET 的下降；在另外的实验中，结合在 DNA 上的荧光探针位于 3′端或下游末端（前缘 FRET），便于研究者在聚合酶靠近 DNA 下游末端时测定 FRET 的上升。图 6.13 说明了后缘荧光共振能量转移和前缘荧光共振能量转移的实验策略。

后缘 FRET 策略不能区分启动子清除后的 σ 因子释放模型和非释放模型。在这两种情况下，σ 因子上的荧光供体探针在启动子清除后都逐渐远离位于 DNA 上游末端的荧光受体探针，因而 FRET 下降。事实也是如此，如图 6.14a 所示，当 DNA 上的探针位于上游末端时，FRET 随着时间的延长而降低。

此外，前缘 FRET 策略能够区分上述两种模型（图 6.13b）。如果 σ 因子与核心酶解离，则 FRET 就会降低，如同后缘 FRET 的实验结果。但是，如果 σ 因子不与核心酶解离，那么随着时间的延长 σ 因子逐渐靠近位于 DNA 下游末端的探针，FRET 将会升高。图 6.14b 显示 FRET 确实升高了，这一实验结果支持启动子清除后 σ 因子仍然停留在核心酶上的假说。事实上，FRET 的增加量值表明，100% 的复合体在启动子清除后仍然保留其 σ 因子。

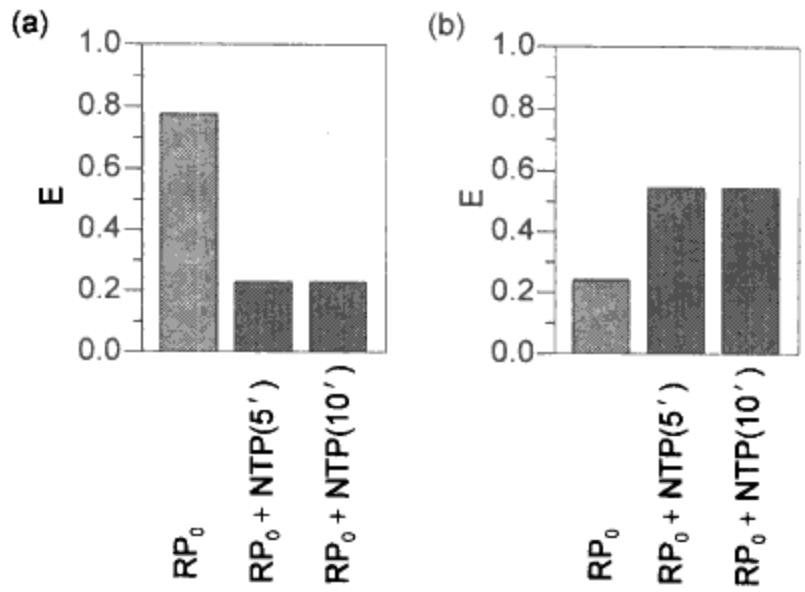


图 6.14 用 FRET 分析启动子清除后 σ 因子与核心酶的结合。Ebright 及同事按图 6.13 描述的方案进行 FRET 分析。（a）后缘 FRET 实验结果；（b）前缘 FRET 实验结果。蓝色柱表示开放启动子复合体（RP_o）的 FRET 效率（E），红色柱分别表示 5min 和 10min 后的 FRET 效率。3 种核苷酸的存在使聚合酶向启动子下游移动 11bp的距离。

Ebright 及同事在聚丙烯酰胺凝胶上完成了图 6.14a 和图 6.14b 所示的实验。研究者首先使开放启动子复合体在溶液中形成，随后加入肝素，使之与游离的聚合酶结合，最后将复合体在聚丙烯酰胺凝胶上进行非变性凝胶电泳。电泳后，在凝胶上标出复合体所在的位置，切胶，将含有复合体的胶条放入荧光测量仪（荧光计）的比色杯中，加入转录缓冲液，测量 RP_o 的 FRET 效率。然后再向比色杯中加入 3 种核苷酸，使 RNA 聚合酶向下游移动，测量延伸复合体的 FRET 效率。这种胶内（in-gel）分析法的优点是只测量活性复合体的 FRET 效率，因为通过凝胶电泳可以除去非活性（闭合启动子）复合体。为消除因电泳而引起可能人为干扰，Ebright 及同事用溶液做了相同的实验，并获得了十分类似的结果。

如果 σ 因子未与核心酶解离，那么如何解释 Travers 和 Burgess 的实验所显示的 σ 因子可被不同核心酶重复利用的结果呢？σ 因子确实可以被重复利用，尽管在延伸过程中 σ 因子仍停留在核心酶上，只是由原来起始阶段的紧密结合状态转变为延伸阶段的松散结合。因此若延伸停止，则 σ 因子就会被

招募到其他核心酶上。在 Ebright 及同事的实验中不会发生这种情况，因为研究者在聚合酶与 DNA 结合后向反应体系中加入了肝素，使所有游离的全酶或核心酶均与肝素结合。

如此，或许存在 σ 因子循环，但这种循环是指 σ 因子与核心酶在起始阶段时的紧密结合与在延伸阶段时的松散结合之间的循环。2001 年，Bar-Nahum 和 Nudler 为这一模型提供了证据。研究者使全酶与带有启动子的 DNA 形成复合体，然后加入 4 种核苷酸中的 3 种，以便聚合酶能够移动到＋32 位置处，然后快速温和地纯化延伸复合体（称为 EC32），即让延伸的 RNA 链的上游末端与束缚在树脂珠（resin bead）上的互补寡核苷酸温和退火，然后低速离心即可快速纯化带有复合物的树脂珠。这种方法只纯化延伸复合体，因为只有延伸复合体携带初生 RNA，可与互补的寡核苷酸链结合。

最后，Bar-Nahum 和 Nudler 用核酸酶处理使复合体与树脂珠分离，对蛋白质做 SDS-PAGE，进行免疫印迹（第 5 章），鉴别复合体中的蛋白质组分。图 6.15 显示纯化的 EC32 复合体至少包含一些 σ 因子。定量分析结果显示，从静止期细胞分离的复合体中，含有 σ 因子的占（33±2)%；从指数生长期细胞分离出的复合体中，仅有（6±1)%的复合体结合有 σ 因子，远低于 Ebright 及同事观察到的 100%。这暗示在延伸复合体中，σ 因子与核心酶间的结合相对较弱。然而，即使保留 σ 因子的复合体数量只有这么多也会显著地促进转录的再起始，因为 σ 因子与核心酶的结合是转录起始的限速步骤。

小结 在延伸阶段，σ 因子能够改变其与核心酶的关系，但是可能并未与核心酶解离，而是转换位置，使之与核心酶间的结合变得十分松散。

启动子处 DNA 的局部熔解 Chamberlin 对 RNA 聚合酶与启动子间相互作用的研究表明，提高温度可使 RNA 聚合酶-启动子复合体变得非常稳定，暗示 DNA 在与聚合酶紧密结合部位发生了局部解链，因为高温能够稳定解链的 DNA。

1978 年，Tao-shih Hsieh 和 James Wang 为 DNA 的局部解链提供了更为直接的证据。研究者将 *E. coli* RNA 聚合酶与含有 3 个 T7 噬菌体菌体早期启动子的限制酶酶切片段结合，测定由此引起的增色效应（第 2 章）。在 260nm 处 DNA 吸光度的增加不仅表明 DNA 双链发生了分离，而且与解链的碱基对数目直接相关。在已知与 DNA 结合的 RNA 聚合酶全酶数量的前提下，Hsieh 和 Wang 计算出每个聚合酶可引起 10bp 的解离。

1979 年，Ulrich Siebenlist 鉴定出 T7 噬菌体菌体早期启动子因 RNA 聚合酶结合而解链的碱基对。

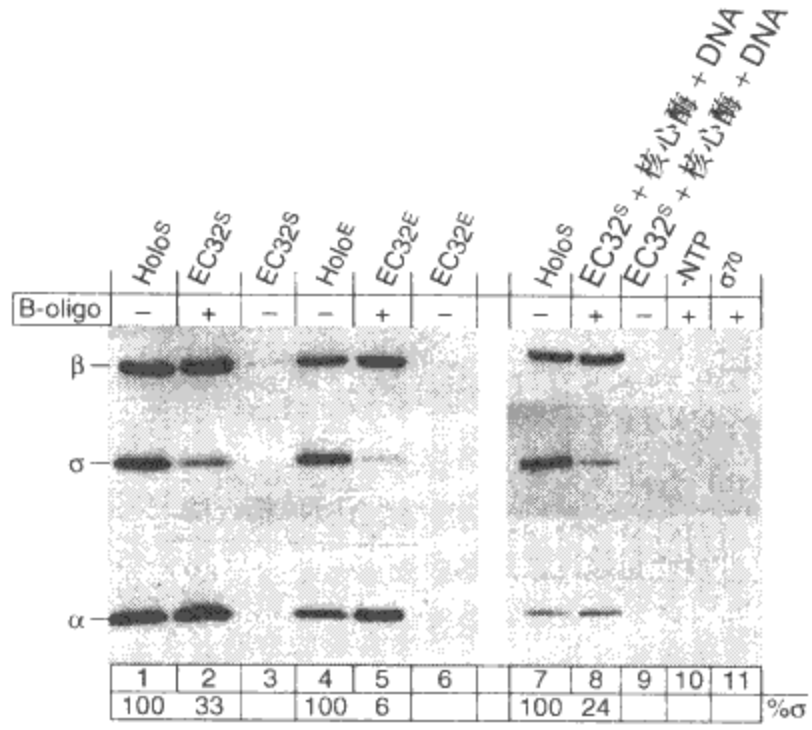


图 6.15 σ 因子与转录延伸复合体结合的测定。 Bar-Nahum 和 Nudler 分别从静止期细胞或指数生长期细胞中纯化出停留在＋32 位置处的延伸复合体（分别表示为 $EC32^S$ 复合体和 $EC32^E$ 复合体），用核酸酶消化新生 RNA 而纯化蛋白质，进行 SDS-PAG 和免疫印迹。复合体的种类以及用于纯化复合体的树脂珠（吸附有寡核苷酸或无核苷酸）在图上方标出。泳道 8 和 9 为对照，即 $EC32^S$ 复合体在与树脂珠结合前先用过量的核心酶和 DNA 处理，目的是排除因 σ 因子与核心酶或 DNA 间的非特异性结合而附着在树脂珠上。（*Source*: Reprinted from *Cell* v. 106, Bar-Nahum and Nudler, p. 444 © 2001, with permission from Elsevier Science.）

图 6.16 为研究者的实验策略。首先，对启动子 DNA 进行末端标记，随后加入 RNA 聚合酶，形成开放启动子复合体。正如我们已经了解到的那样，开放启动子复合体的形成涉及局部 DNA 的解链。当双链分离时，A 的 N_1（正常情况下与互补链的 T 通过氢键配对）易受到某些化学试剂的攻击。Siebenlist 用硫酸二甲酯（dimethyl sulfate，DMS）对暴露的腺嘌呤进行甲基化修饰，当去除 RNA 聚合酶后，解链区可重新缔合，但是甲基基团阻止了 N_1－甲基－腺嘌呤与互补链的胸腺嘧啶间的正确碱基配对，这样至少保留了原先解链区的部分单链特征。然后用特异性切割单链 DNA 的 S1 核酸酶消化 DNA。该酶会在启动子解链区内被甲基化修饰的腺苷酸处切割，产生一系列末端标记片段，每个片段的末端均为解链区的腺嘌呤。最后，电泳标记的 DNA 片段以确定其准确长度。根据片段长度和末端标记的确切位置，就能精确计算出了解链区的位置。

图 6.17 是实验结果。我们获得的不是预期的整齐片段，而是从＋3 到－9 大小不等的模糊条带。原

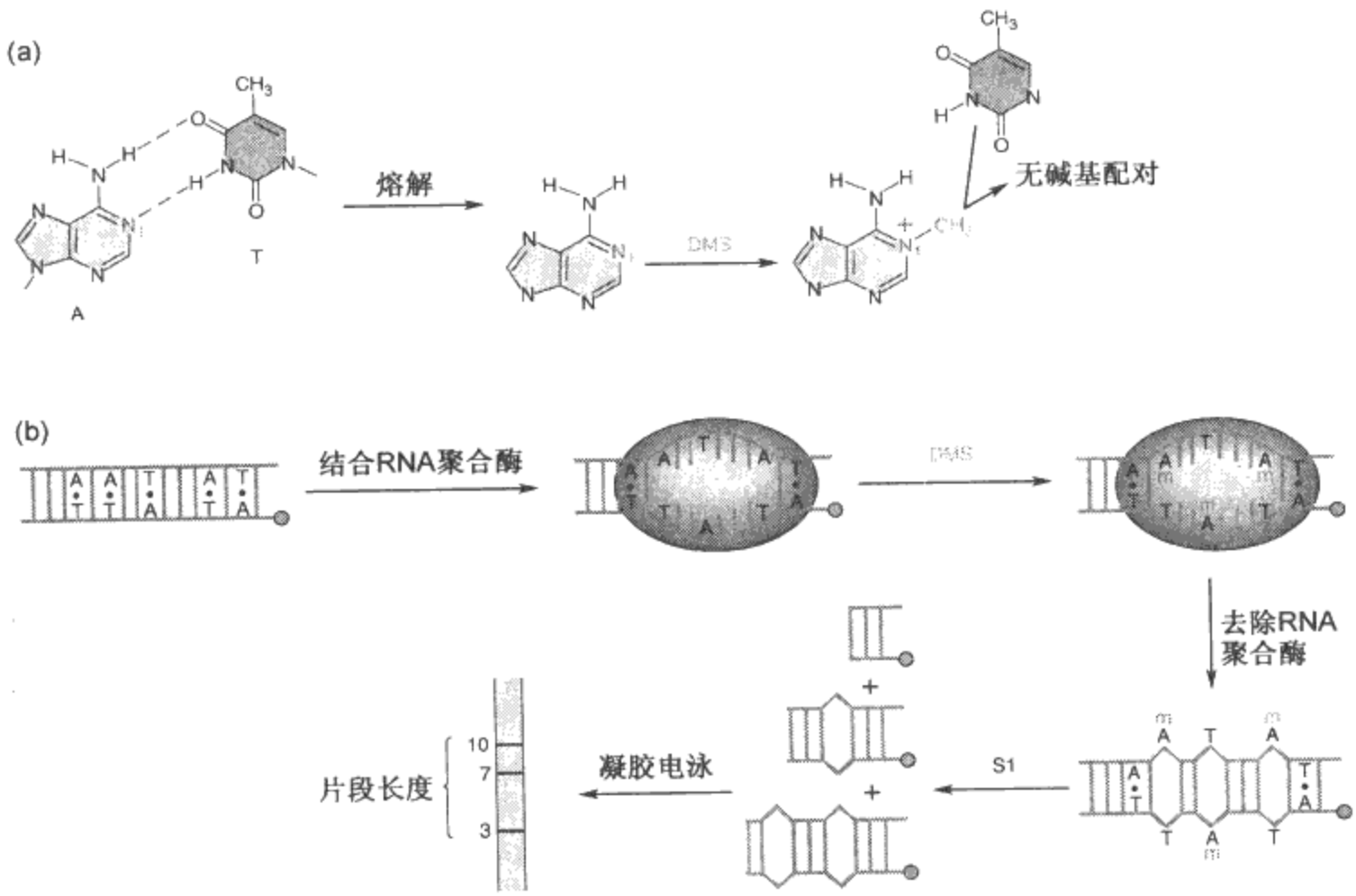


图 6.16　RNA 聚合酶引发 T7 噬菌体早期启动子熔解区域的定位。(a) 当腺嘌呤与胸腺嘧啶（左）形成碱基对时，腺嘌呤的 N_1 隐藏在双螺旋的中间而不能被甲基化修饰。当发生熔解时，A 与 T 分离，此时，A 易受到硫酸二甲酯（DMS，蓝色）的攻击，N_1 发生甲基化修饰。被甲基化修饰了的 A 不再与 T 互补配对。(b) 假设包含有 5 个 A-T 碱基对的启动子区域末端被标记（橙红色）。RNA 聚合酶（红色）的结合会使启动子 DNA 序列发生局部熔解，3 个 A 被甲基化修饰，当聚合酶脱离后，A 失去与 T 重新配对的能力。在十分温和的条件下，用 S1 核酶在单链区域切割 DNA，每次只切割一个 DNA 分子，产生短片段，变性后电泳即可判断其大小。片段的长度可以表明 DNA 的解链区域与标记末端的距离。

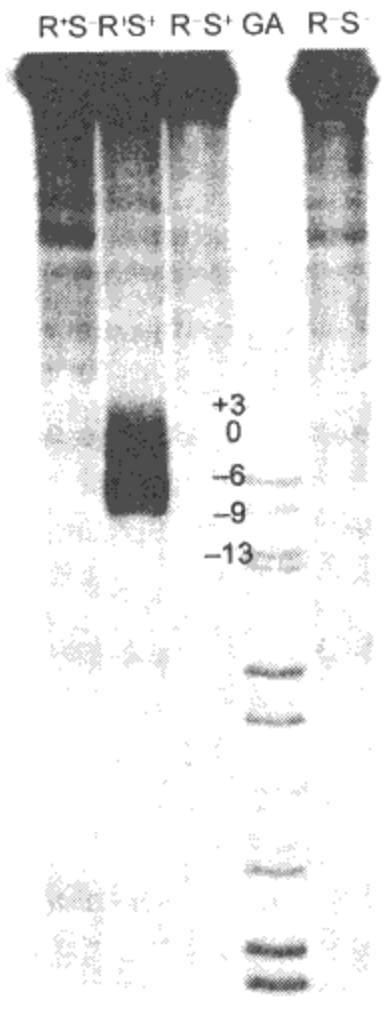
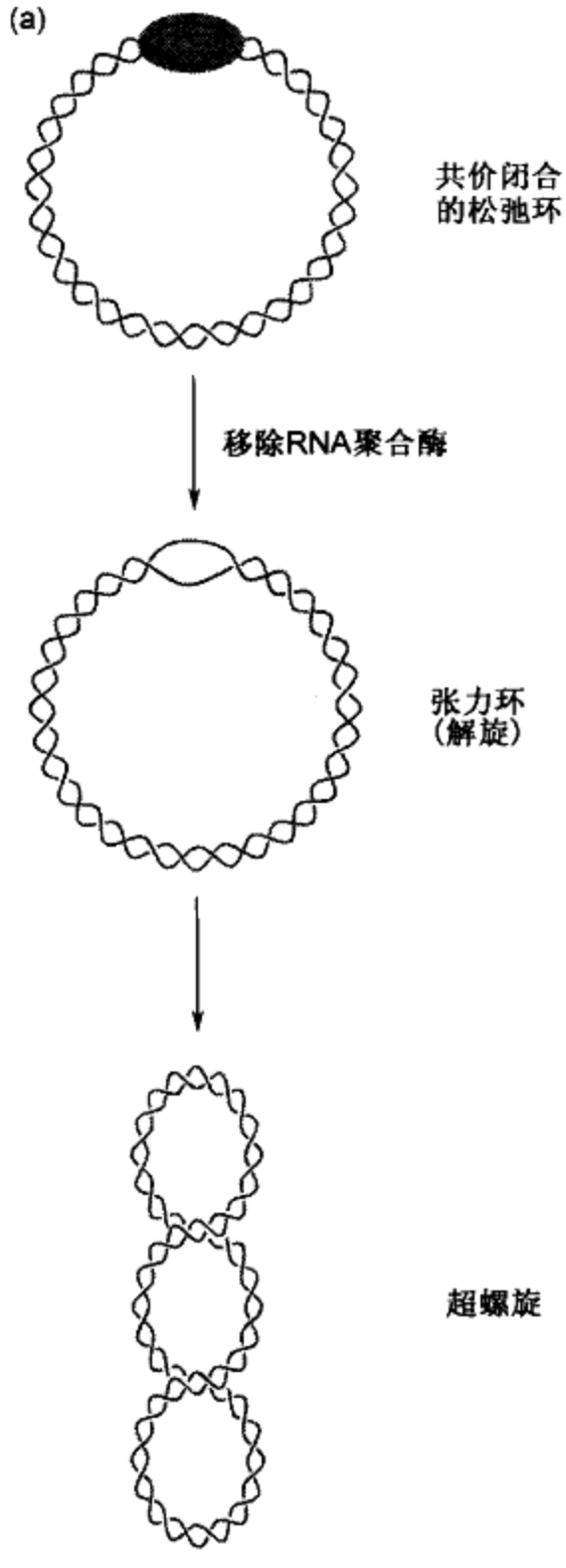


图 6.17　RNA 聚合酶使 T7 噬菌体 A3 启动子 DNA 在－9 至＋3 区域发生熔解。Siebenlist 进行了图 6.16 所示的甲基化-S1 核酸分析。泳道 R^+S^+ 为 RNA 聚合酶（R）和 S1 核酸酶（S）同时存在时的实验结果，其他泳道为对照，即无 RNA 聚合酶或 S1 核酸酶，或者两者均无。泳道 GA 是启动子部分片段测序结果，使 Siebenlist 将启动子的解链区域定位在－9 至＋3 区间内。(*Source*: Siebenlist. RNA polymerase unwinds an 11-base pair segment of a phage T7 promoter. *Nature* 279 (14 June 1979) p. 652, f. 2, © Macmillan Magazines Ltd.)

因可能是解链区的多甲基化修饰引入了正电荷，使碱基配对能力大为减弱，以至于不能形成强碱基对。整个解链区域至少部分保留了单链特征，从而可被 S1 核酸酶切割。实验测定的解链区长度为 12bp，大体上与 Hsieh 和 Wang 的估算值一致，这可能是个低估值，因为位于酶解片段两端、与 A-T 碱基对相邻的 G-C 未被检出，因为在该实验中鸟嘌呤和胞嘧啶都不能被甲基化修饰。这一结果也符合解链区域恰是 RNA 聚合酶起始转录位点的论断。

Hsieh 和 Wang、Siebenlist 的实验以及其他前期实验均检测了在 RNA 聚合酶与 DNA 形成的简单二元复合体中 DNA 的解链情况。未测定 DNA“熔解泡”的大小，而 DNA“熔解泡”恰是 RNA 链合成起始与延伸的位点。为此，1982 年 Howard Gamper 和 John Hearst 着手研究因 RNA 聚合酶结合而熔解的碱基对数目，在二元复合体以及包含 RNA 的活性转录复合体（三元复合体）中，研究者以 SV40 DNA 为研究对象，该 DNA 具有一个能被 *E. coli* RNA 聚合酶识别的启动子。在 5℃或 37℃及核苷酸缺乏条件下将 RNA 聚合酶与 SV40 DNA 结合形成二元复合体，或者添加核苷酸形成三元复合体。实验中每个 RNA 聚合酶仅起始转录一次且不终止转录。所有的聚合酶始终与 DNA 结合，这样就能准确估算与 DNA 结合的 RNA 聚合酶的数目。

在已知与 DNA 结合的 *E. coli* RNA 聚合酶数目后，Gamper 和 Hearst 松弛与人细胞粗提物形成的任何超螺旋，然后从松弛的 DNA 上去除聚合酶（图 6.18a），蛋白质去除后留下 DNA 的解链区，这意味着整个 DNA 处于解链状态。由于 DNA 仍然处于共价闭合的环形结构，所以由解旋而引入环内的张力需通过形成超螺旋而释放。超螺旋程度越多，RNA 聚合酶引起双螺旋解旋区域就越大。DNA 超螺旋程度可以通过凝胶电泳加以测定，DNA 超螺旋数越多，其电泳迁移速度就越快。



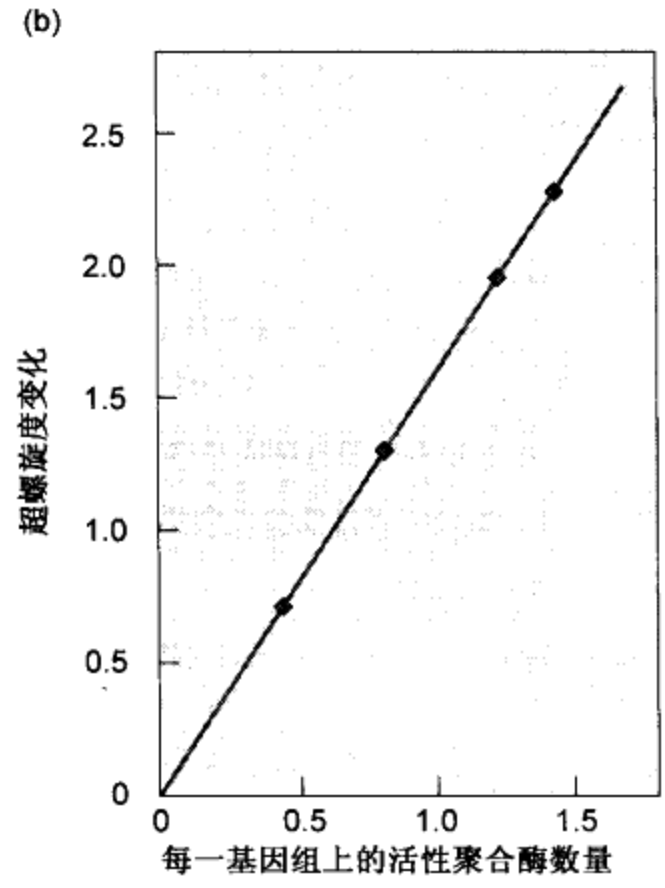


图 6.18　RNA 聚合酶结引起 DNA 解旋区长度的测定。（a）实验原理。Gamper 和 Hearst 将 *E. coil* RNA 聚合酶（红色）加入到 SV40 DNA 中，用切口-闭合（nicking-closing）提取物松弛超螺旋，产生如图上部所示的复合体。然后去除聚合酶，因为该区已被聚合酶解链，所以留下具有张力的 DNA（图中间所示），这个张力通过形成超螺旋而迅速释放（图底部所示）。超螺旋程度越高，RNA 聚合酶引起的解链区长度就越大。（b）实验结果。Gamper 和 Hearst 根据 RNA 聚合酶与超螺旋度之间函数关系绘制变化曲线，该直线的斜率约为 1.6（每个聚合酶可引入 1.6 个超螺旋）。

图 6.18b 是在 37℃条件下超螺旋的变化图示，超螺旋度是每个基因组中活性聚合酶数目的函数。这两个变量间存在着线性关系，每个聚合酶可引起约 1.6 个超螺旋，即一个聚合酶解旋 1.6 圈的 DNA 双螺旋。如果每圈 DNA 双螺旋中含有 10.5bp，那么每个 RNA 聚合酶解链的长度约 17bp（1.6×10.5=16.8）。分析 5℃条件下的实验数据得出了类似结论，即每个 RNA 聚合酶解链 DNA 的长度为 18bp。根据这些数据，Gamper 和 Hearst 得出结论：RNA 聚合酶与启动子结合后，会引起 DNA 局部解链而形成长度约为（17±1)bp 的转录泡（transcription bubble)。随着转录的进行，转录泡随聚合酶移动。当然，DNA 解链是必需的，因为通过解旋使模板链暴露出来而得以转录。

小结 与启动子结合后，RNA 聚合酶会引起邻近转录起始位点处 DNA 的熔解，解链的长度为10～17bp。转录泡随聚合酶的移动而移动，暴露出模板链以便转录。

σ 因子的结构

在 20 世纪 80 年代后期，从多种细菌中克隆并测序了编码各种 σ 因子的基因。在第 8 章我们将了解到每类细菌都有一种基本 σ 因子（primary σ-factor)，负责转录营养生长阶段的基因，这些基因都是细菌日常生长所必需的，如 *E. coli* 的 σ^{70}、枯草芽孢杆菌 *B. subtili* 的 σ^{43}。名称是根据各自的分子质量分别为 70kDa 和 43kDa 命名的。根据其主要特性又将这些蛋白质命名为 σ^{A}。此外，细菌还拥有用于转录特异基因（如热击基因、芽孢基因等）的选择性 σ 因子。1988 年，Helmann 和 Chamberlin 综述，并分析了文献中的所有这类因子，发现其氨基酸序列具有惊人的相似性，具有串联成簇的 4 个区域（区 1 至区 4，见图 6.19)。这些区域的序列保守性表明它们对 σ 因子的功能十分重要。事实上，这些区域均参与对核心酶的结合，并且对核心酶与 DNA 的结合起正调控或负调控的作用。Helmann 和 Chamberlin 推测了各区域可能具有的功能。

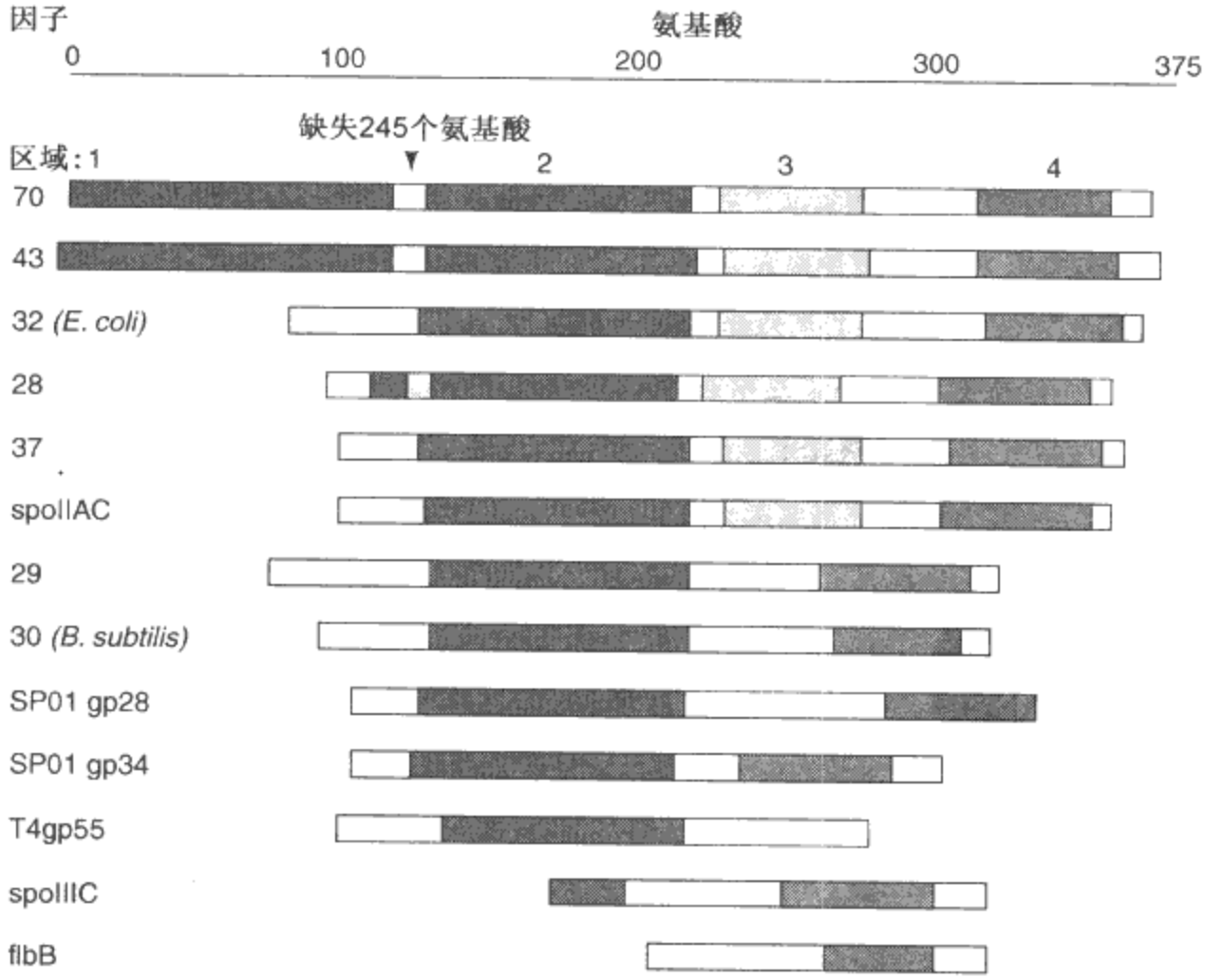


图 6.19 ***E. coli* 和 *B. subtilis* 各种 σ 因子的同源区。**σ 蛋白用水平线条带表示，垂直排列的为同源区。顶部的两条带分别代表 *E. coli* 与 *B. subtilis* 的基本 σ 因子 σ^{70} 和 σ^{43} 均含有同源区 1，另外，在区 1 与区 2 之间，σ^{70} 含有一段由 254 个氨基酸残基组成的序列，而 σ^{43} 则无这一序列，该序列标注在 σ^{70} 条带的上方。浅阴影部分表示仅在某些蛋白质中存在的保守区。

区域 1 该区仅被发现存在于基本 σ 因子（σ^{70} 和 σ^{43}）中，其功能是阻止 σ 因子自身直接与 DNA 结合。在本章后面部分我们将了解到，σ 因子的一个片段能与 DNA 结合，但 1 区可阻止整个多肽与 DNA 的结合。这对转录事件来讲是重要的，因为 σ 因子与启动子结合会抑制全酶的结合，进而抑制转录。

区域 2 该区存在于所有的 σ 因子中，是 σ 因子最保守的区域，可分为 2.1～2.4 区 4 个部分（图 6.20）。

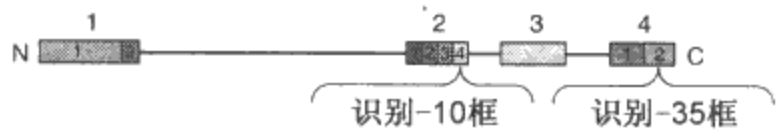


图 6.20 *E. coil* σ^{70} 的基本结构区域。图中显示了 4 个保守区及 1、2 和 4 的亚区。（*Source*: Adapted from Dombroski, A. J., et al., "Polypeptides containing highly conserved regions of the transcription initiation factor σ^{70} exhibit specificity of binding to promoter DNA." *Cell* 70: 501-12, 1992.）

有充分证据表明，2.4 区负责 σ 因子识别启动子－10 框的关键活性。首先，如果 2.4 区确实识别－10 框，那么具有相似专一性的 σ 因子都应该具有相似的 2.4 区。事实确实如此；*B. subtilis* 的 σ^{43} 和 *E. coli* 的 σ^{70} 能够识别相同的启动子序列，包括－10 框。事实上，这两个 σ 因子可以互换，其 2.4 区有 95%的相似性。

Richard Losick 及同事通过遗传实验使 2.4 区与－10 框连接，σ 因子的 2.4 区有一段氨基酸序列可形成 α 螺旋，在第 9 章将学习 α 螺旋是理想的 DNA 结合基序，这与 σ 因子的 2.4 区参与启动子结合的作用是一致的。Losick 及同事分析认为，如果该潜在的 α 螺旋是真正的－10 框识别区，那么可以开展如下实验。首先，改变启动子－10 框的单个碱基，破坏－10 框与 RNA 聚合酶的结合。然后在 σ 因子的 2.4 区进行某个氨基酸的补偿突变。如果 σ 因子的这种补偿突变能够抑制启动子突变，恢复对突变启动子的结合，这就为－10 框与 σ 因子的 2.4 区确实关联提供了强有力的证据。于是，Losick 及同事对 *B. subtilis* 的 *spo*VG 启动子－10 框进行了 G→A 的转换突变，破坏启动子与聚合酶的结合。然后对 σ^{H}2.4 区第 100 位氨基酸进行了 Thr→Ile 突变，结果恢复了聚合酶识别突变启动子的能力。

区域 3 参与核心酶与 DNA 的结合，相关内容将在本章后面学习。

区域 4 与区域 2 相似，区域 4 也可被分为几个亚区，在启动子识别过程中也起关键作用。4.2 亚区有一个螺旋-转角-螺旋的 DNA 结合域（第 9 章），表明该区在聚合酶-DNA 结合中起作用。事实上，4.2 亚区有可能控制 σ 因子与启动子－35 框的结合，如同 σ2.4 区与启动子－10 框的关系。遗传学和其他证据都支持 4.2 亚区与－35 框的关联。而且我们观察到，识别相似－35 框启动子的 σ 因子具有相似的 4.2 亚区。此外，σ 因子 4.2 亚区的补偿突变可抑制启动子（在－35 框）的突变。例如，Miriam Susskind 及同事证明 *E. coli* σ^{70} 的第 588 位 Arg→His 突变抑制了 *lac* 启动子－35 框序列 G→A 或 G→C 的突变。图 6.21 总结了 σ 因子 2.4 亚区和 4.2 亚区分别与细菌启动子－10 框及－35 框的相互作用。

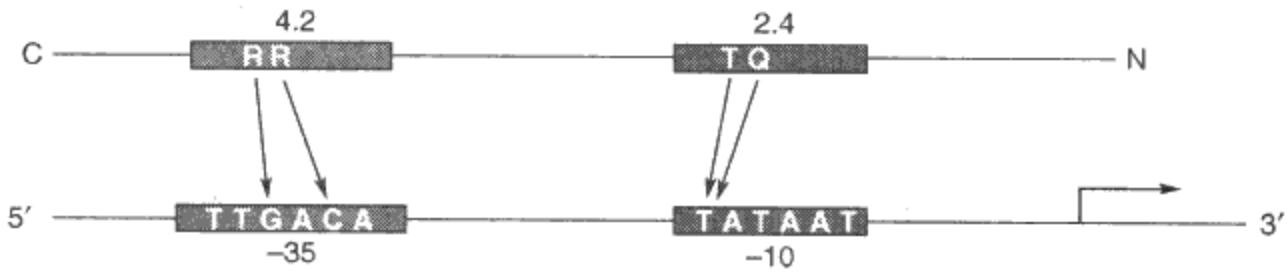


图 6.21 σ 与启动子不同区域间的特异性相互作用。箭头表示 σ^{70} 抑制突变实验揭示出的 σ 因子与启动子间特异性相互作用关系，上方线条内的字母代表 σ^{70} 蛋白中突变的氨基酸残基，箭头所指启动子中的碱基是与 σ^{70} 蛋白中相应氨基酸直接发生接触的碱基。σ^{70} 蛋白 4.2 亚区中的 2 个 R 代表第 584 位和第 588 位的精氨酸残基，分别与启动子－35 框的 C 及 G 发生特异性接触。σ^{70} 蛋白 2.4 亚区中的 T 和 Q 分别代表第 437 位的谷氨酸残基和第 440 位的苏氨酸残基，均能与启动子－10 框的 T 发生特异性接触。值得注意的是，σ 因子线性结构（上方线条）的书写方式是 C 端在左侧，与启动子序列相匹配，启动子序列的习惯书写方式是从左至右，即 5′→3′（下方线条）。（*Source*: Adapted from Dombroski, A. J., et al., "Polypeptides containing highly conserved regions of the transcription initiation factor σ^{70} exhibit specificity of binding to promoter DNA." *Cell* 70: 501-12, 1992.）

这些研究结果均表明了2.4亚区和4.2亚区在σ因子分别与启动子－10框和－35框结合过程中的重要性。从其重要地位来看，σ因子甚至还应具有DNA结合域，但令我们困惑的事实是，σ因子自身并不与启动子或DNA的其他任何区域结合，只有当σ因子与核心酶结合后才能与启动子结合。我们如何解释这个显而易见的矛盾呢？

Carol Gross及同事认为σ因子的2.4亚区和4.2亚区本身都能与相应的启动子区结合，但是σ因子的其他结构域干扰了这种结合。现在我们知道，1.1区在没有核心酶时会阻止σ因子与DNA结合。研究者进一步指出，当σ因子与核心酶结合后其自身构象发生改变，暴露出其DNA结合域，从而使σ因子与启动子结合。为了验证这一假设，研究者构建了谷胱甘肽-S-转移酶（glutathione-S-transferase，GST）和*E. coli* σ因子片段（包括2.4亚区或4.2亚区，或这两个区域均包含）的融合蛋白（第4章），利用GST对谷胱甘肽的亲和性可方便地纯化融合蛋白。结果表明，2.4亚区融合蛋白能够与含有－10框的DNA片段结合，但不能结合含有－35框的DNA片段。4.2亚区融合蛋白能够与含有－35框而不是－10框的DNA片段结合。

为检测融合蛋白与启动子元件的结合情况，Carol Gross及同事采用了硝化纤维素滤膜结合分析策略。研究者先标记含有一个或两个*tac*启动子元件的靶DNA，*tac*启动子含有*lac*启动子的－10框及*trpc*启动子的－35框。然后在过量未标记竞争DNA存在的情况下，将融合蛋白加入到被标记的靶DNA中，通过硝化纤维素结合实验测定形成的标记DNA-蛋白质复合体。

图6.22a是Gross及同事所做实验的一个结果，标记的*tac*启动子可与GST-σ-4区融合蛋白结合，因为σ因子4区含有假定的－35框结合域。我们预期该融合蛋白与含有*tac*启动子的DNA的结合能力要强于与不含*tac*启动子的DNA的结合能力，图6.22a显示的结果恰是如此。含有*tac*启动子的未标记DNA是理想的竞争者，而缺失*tac*启动子的未标记DNA的竞争力则相对较弱。这样，GST-σ-4区融合蛋白与非特异性DNA的结合力较弱，而与含有*tac*启动子DNA片段的结合力较强，正是我们预期的结果。

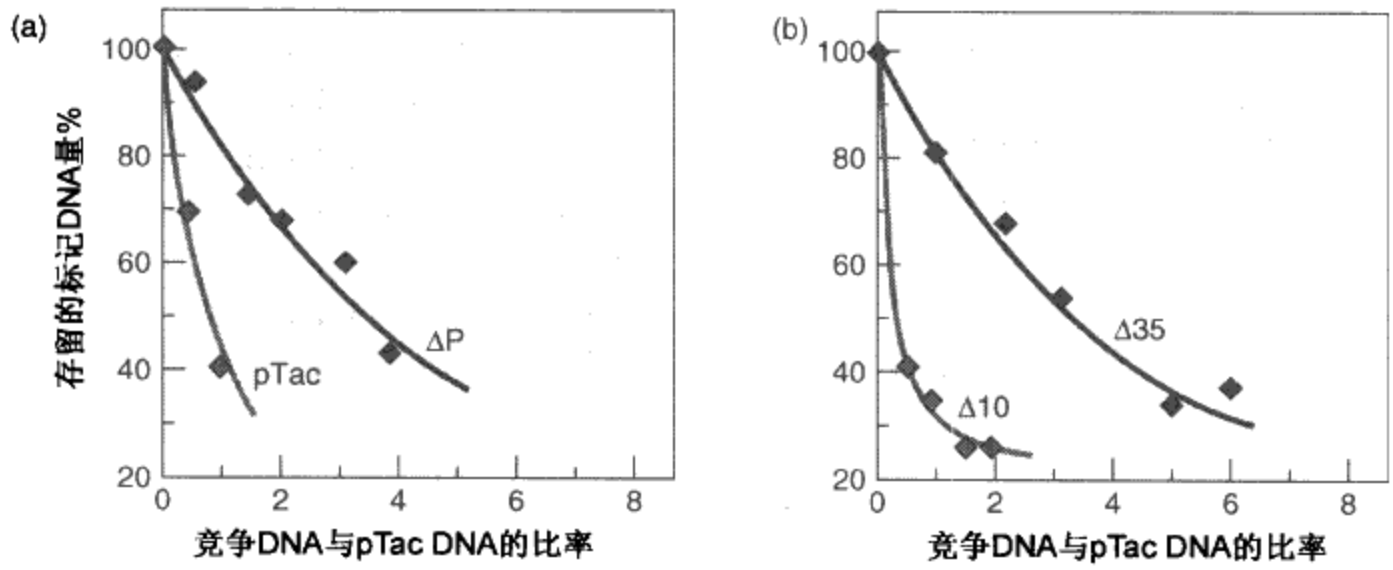


图6.22　σ因子4.2区与启动子－35框之间的结合分析。（a）启动子的识别。Gross及同事分析σ-GST融合蛋白与含*tac*启动子的标记DNA片段（pTac）的结合。实验中使用的σ片段是*E. coli* σ因子C端的108个氨基酸序列，只有4区而无2区。在竞争者为pTac DNA（含有*tac*启动子）或ΔP DNA（不含有*tac*启动子）的情况下，测定标记的DNA-蛋白质复合体与硝化纤维素膜的结合。由于pTac DNA比ΔP DNA具有更强的竞争力，因此推断含4区的融合蛋白能结合*tac*启动子。（b）－35框的识别。Gross及同事重复上述试验，只是竞争DNA不同：Δ10在*tac*启动子－10框缺失6bp；Δ35在*tac*启动子－35框缺失6bp。缺失－35框与无*tac*启动子相比根本没有增加竞争力，而缺失－10框没有降低竞争力，说明σ片段通过4区结合－35框而不是－10框。（*Source*：Adapted from Dombroski，A. J.，et al.，"Polypeptides containing highly conserved regions of transcription initiation factor σ^{70} exhibit specificity of binding to promoter DNA." *Cell* 70：501-12，1992.）

图6.22b显示GST-σ-4区融合蛋白与含－35框而不含－10框启动子结合。正如所见，删除－35框的竞争DNA并不比非特异性DNA更具竞争力，而删除－10框的竞争DNA因还保留－35框而具有更强的竞争力。因此，σ因子4区能特异性地结合－35框而非－10框。用GST-σ-2区融合蛋白进行类

似实验表明，该融合蛋白能特异性地结合－10 框而非－35 框。

我们已经知道，聚合酶全酶能够识别启动子，通过使－11 至 ＋1 区域的一小段 DNA 双链熔解而形成开放启动子复合体，推测 σ 因子在该过程中起重要作用，但 σ 因子自身并不能形成开放启动子复合体。开放复合体形成的特征之一是聚合酶结合在启动子－10 框的非模板链上。同样，σ 因子不可能单独完成这一过程，推测可能是核心酶的某些片段协助 σ 因子完成了这项工作。Gross 及同事提出了这样一个问题：核心酶的哪一部分是暴露 σ 因子与启动子－10 框非模板链结合的部位所必需的？

为回答这个问题，Gross 及同事将注意力集中在 β′亚基上，已经证明该亚基能与 σ 因子协同与－10 框非模板链结合。研究者克隆了 β′亚基的不同片段，然后同 σ 因子一起检测与放射性标记的单链寡核苷酸的结合能力，这些单链寡核苷酸相应于启动子－10 框的模板链或非模板链。将 β′片段、σ 因子和标记 DNA 共同孵育，然后将混合物经紫外交联仪照射使 σ 因子与 DNA 交联。对交联复合物进行 SDS-PAGE 分析。如果 β′片段能够诱导 σ 因子与 DNA 的结合，那么，σ 因子就会与标记 DNA 交联，对应于 σ 因子的 SDS-PAGE 条带就会被标记。

图 6.23 显示了包含 1～550 个氨基酸残基的 β′片段可引起 σ 因子与非模板链 DNA 的结合（但不是模板链），而 σ 自身则几乎无 DNA 结合能力。接下来，Gross 及同事用 β′亚基 1～550 区中更小的片段查验 β′亚基诱导结合的具体部位。图 6.23 显示所有片段均能诱导 σ 因子与 DNA 结合，尽管 260～550 片段只在低温下起作用。特别是 262～309 这一个只有 48 个氨基酸残基的小片段即使在室温下也具有很强地促进结合活性，已知该区的 R275、E295 和 A302 这 3 个氨基酸残基突变会干扰 σ 因子与 DNA 的结合。相应地，Gross 及同事检验了这些突变干扰 σ 因子与－10 框非模板链结合的情况，结果表明，每一突变都会产生很强的干扰作用。

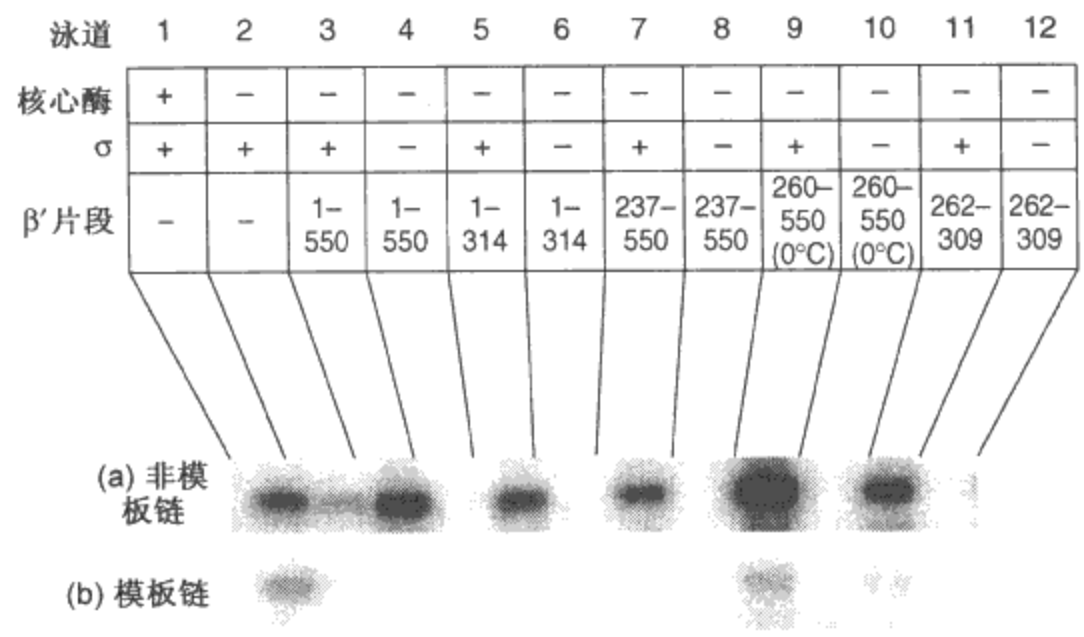


图 6.23　σ 因子与启动子－10 框的结合诱导。Gross 及同事将 σ 亚基与长度不一的 β′片段混合，如图上方所示，然后加入分别代表启动子－10 框模板链或非模板链的标记寡核苷酸单链，经紫外线照射使 σ 因子与 DNA 交联，然后对交联复合体进行 SDS-PAGE 分析，并放射自显影检测结合标记 DNA 的 σ 因子。泳道 1：完整核心酶的阳性对照；泳道 2：无 β′片段的对照；其他所有偶数泳道为无蛋白的阴性对照。泳道 9 和 10 是在 0℃下进行的；其他几组实验均在室温下进行。（a）非模板链；（b）模板链。（*Source*: Reprinted from *Cell* v. 105, Young et al., p. 940 © 2001, with permission from Elsevier Science.）

小结　不同 σ 基因的序列比对结果显示，不同种类的 σ 因子有 4 个相似的区域。2.4 亚区和 4.2 亚区分别识别启动子－10 框及－35 框。σ 因子自身并不能与 DNA 结合，但与核心酶相互作用后暴露出 σ 因子的 DNA 结合域。具体讲，就是 β′亚基 262～309 间的氨基酸片段促进 σ 因子与启动子－10 框非模板链的结合。

α 亚基在 UP 元件识别中的作用

本章前面部分讲到，RNA 聚合酶自身能识别称为 UP 元件的上游启动子元件，我们知道 σ 因子识别核心启动子元件，但聚合酶的哪个亚基负责对 UP 元件的识别呢？基于以下证据，显然是核心酶的 α 亚基。

Richard Gourse 及同事构建了 α 亚基发生不同

突变的 *E. coli* 菌株，发现其中有些突变体失去应答 UP 元件的能力，因为这些突变体从有 UP 元件的启动子处起始的转录并不比从无 UP 元件的启动子处起始的转录水平高。为测定转录水平，研究者将强野生型启动子 *rrnB* P1 或缺失 UP 元件的突变型启动子置于克隆载体的 *rrnB* P1 转录终止子上游 170bp 处。用 3 种不同的 RNA 聚合酶转录这些重组子，这 3 种聚合酶由纯化的亚基重组而成：①具有正常 α 亚基的野生型聚合酶；②α-235，α 亚基其 C 端缺失 94 个氨基酸的聚合酶；（3）R265C，α 亚基第 265 位氨基酸发生 Arg→Cys 突变的聚合酶。用标记的核苷酸标记 RNA 产物，然后进行凝胶电泳，放射自显影检测 RNA 产物。

图 6.24a 是野生型聚合酶的转录结果。野生型启动子（泳道 1 和 2）的转录活性远高于 UP 元件被载体 DNA 取代的启动子（泳道 3 和 4）或缺失 UP 元件的启动子（泳道 5 和 6）。图 6.24b 是 α 亚基 C 端缺失 94 个氨基酸的聚合酶转录结果。在转录有核心启动子基因时，该聚合酶与野生型聚合酶具有相同的活性（比较图 a 和图 b 中的泳道 3～6），但是与野生型聚合酶相比，突变型聚合酶不能区分有或无 UP 元件的启动子（比较泳道 3～6 与泳道 1 和泳道 2）。在此实验中，UP 元件根本不起作用，可见，α 亚基的 C 端可能使聚合酶具有应答 UP 元件的能力。

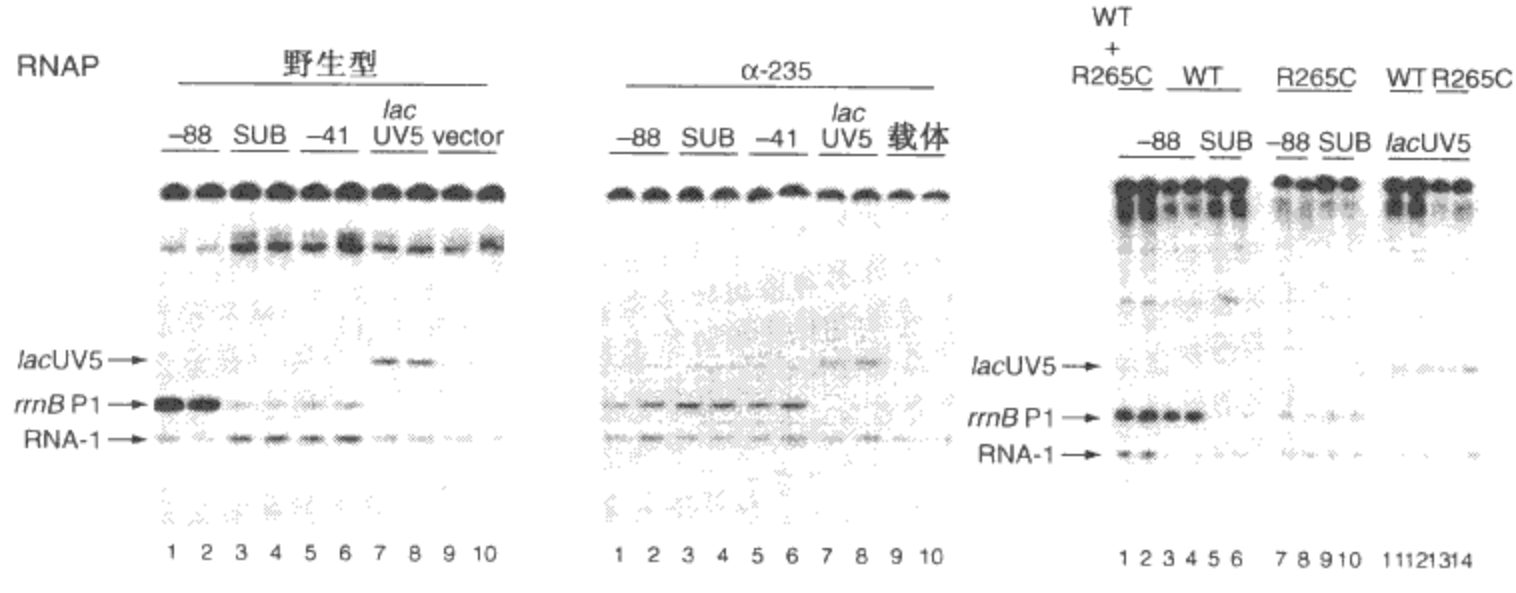


图 6.24　RNA 聚合酶的 α 亚基在 UP 元件识别中的重要性。Gourse 及同事在体外转录含有启动子序列的质粒（图上方所示），研究者将启动子序列插入到转录终止子上游 100～200nt，产生特定长度的转录物。反应结束后，对标记的产物进行琼脂糖凝胶电泳和放射自显影检测，使用的启动子如下：－88，－88～＋1 间是野生型序列；SUB，－59～－41 是无关序列而不是 UP 元件；－41，－41 位上游的 UP 元件被载体序列取代；lacUV5，无 UP 元件的 *lac* 启动子；“载体”是无启动子插入的质粒。转录物的位置在左侧标出，分别是从 rrnBP1 启动子、lacUV5 启动子起始的转录物以及从质粒复制起点起始转录的 RNA-1。图上部的 RNAP 表示试验中所使用的 RNA 聚合酶：（a）野生型聚合酶；（b）α-235 聚合酶（α 亚基的 C 端缺失 94 个氨基酸残基）；（c）野生型聚合酶（WT）或 R265C 聚合酶（C265 替代 R265）。（*Source*: Ross et al., A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science* 262 (26 Nov 1993) f. 2, p. 1408. © AAAS.）

图 6.24c 表明，α 亚基第 265 位正常精氨酸被半胱氨酸取代的聚合酶 R265C 不能应答 UP 元件（泳道 7～10 均显示有适度转录）。因此，该单个氨基酸的改变显然会破坏 α 亚基应答 UP 元件的能力。这一结果不是由 R265C 聚合酶制备过程中抑制剂引起的人为假象，因为 R265C 聚合酶与野生型聚合酶的混合物仍能应答 UP 元件（泳道 1～ 4 均显示强转录）。

为验证 α 亚基确实与 UP 元件结合这一假设，Gourse 及同事用含有 *rrnB* P1 启动子的 DNA 和野生型聚合酶或突变型聚合酶进行 DNase 足迹实验（第 5 章）。他们发现，野生型聚合酶在核心启动子和 UP 元件上都留下了足迹，而 α 亚基 C 端结构域缺失的突变型聚合酶仅在核心启动子上产生足迹（数据未给出），表明 α 亚基 C 端结构域是聚合酶与 UP 元件相互作用所必需的。用纯化的 α 亚基二聚体进行 *rrnB* P1 启动子的 UP 元件足迹实验，Gourse 及同事进一步为这一假设提供了证据。图 6.25 为实验结果，α 亚基二聚体在 UP 元件上留下了清晰的足迹。

Richard Gourse 和 Richard Ebright 及同事通过限制性蛋白酶分析（limited proteolysis analysis）显

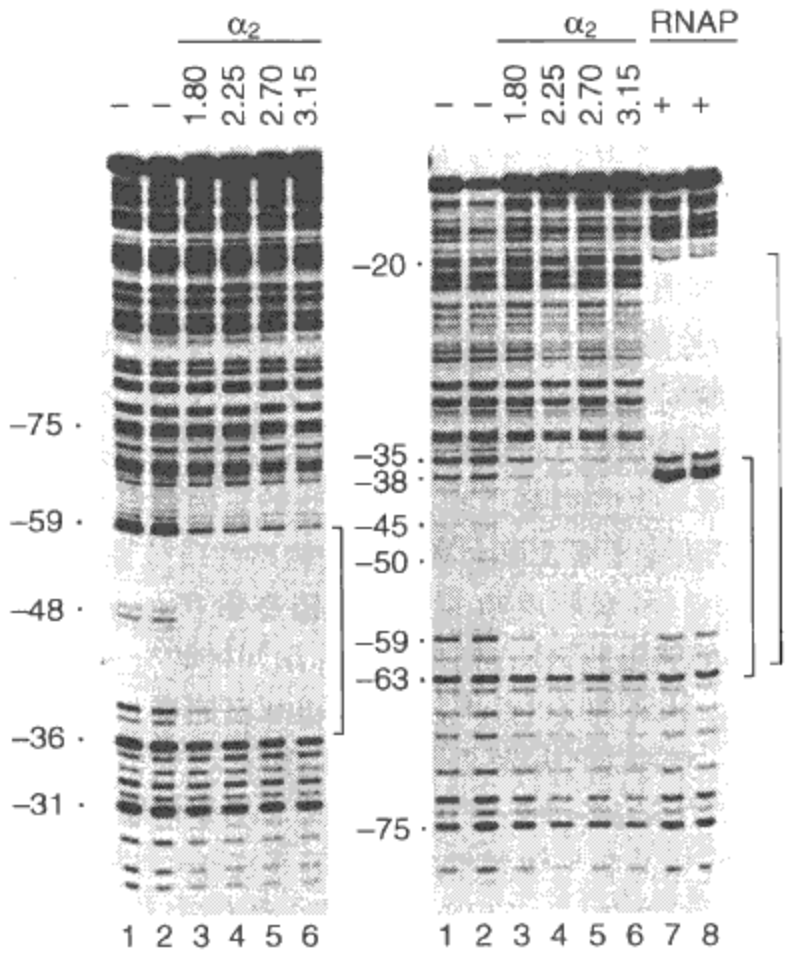


图 6.25　纯化 α 亚基的 UP 元件足迹。Gourse 及同事对 *rrnB* P1 启动子末端标记的模板链（a）和非模板链（b）进行 DNase 足迹实验。纯化 α 亚基二聚体的用量（μg）及 10nmol/L 的 RNA 聚合酶全酶（RNAP）在图上方标出，粗线方括号表示因 α 亚基在 UP 元件中的足迹，细线方括号表示全酶在 UP 元件中的足迹。（*Source*：Ross et al.，A third recognition element in bacterial promoter：DNA binding by the α-subunit of RNA polymerase. *Science* 262（26 Nov 1993）f. 5，p. 1408. © AAAS.）

示 α 亚基的 N 端和 C 端结构域（分别为 α-NTD 和 α-CTD）是独立折叠形成的 2 个结构域，通过柔性接头连接在一起。蛋白质结构域是能够独立折叠而形成特定的结构的蛋白质部分。由于折叠，结构域能抵抗蛋白水解酶。所以蛋白水解酶只能有限消化结构域之间的未折叠部分，留下结构本身。Gourse 和 Ebright 及同事对 *E. coli* RNA 聚合酶的 α 亚基进行限制性蛋白酶水解，获得了一条分子质量约为 28kDa 的多肽和 3 条分子质量约为 8kDa 的多肽。产物末端序列表明，28kDa 多肽包含 8～241 位氨基酸，3 条较小的多肽分别包含 242～329、245～329 和 249～329 位氨基酸。表明 α 亚基折叠形成两个结构域：较大的 N 端结构域，包含第 8～241 位氨基酸；较小的 C 端结构域，包含第 249～329 位氨基酸序列。

此外，这两个结构显然被一个未折叠的接头（linker）所连接，该接头至少有三处能被实验所用的蛋白酶（Glu-C）切割，乍看起来，该接头包括第 242～248 位氨基酸，因为 Glu-C 在切割蛋白质时，要求在其所切割键的两侧均具有 3 个无结构的氨基酸。然而，接头比其最初的表征要长，事实上，至少长度为 13 个氨基酸（239～251 位氨基酸残基）。

基于这些实验提出了图 6.26 所示的模型。RNA 聚合酶通过 σ 因子结合到核心启动子上，不需要 α 亚基 C 端结构域的帮助。但 RNA 聚合酶要与含有 UP 元件的启动子结合，则需要 σ 因子及 α 亚基 C 端结构域的参与，使聚合酶与启动子间产生强烈的相互作用，从而实现高水平的转录。

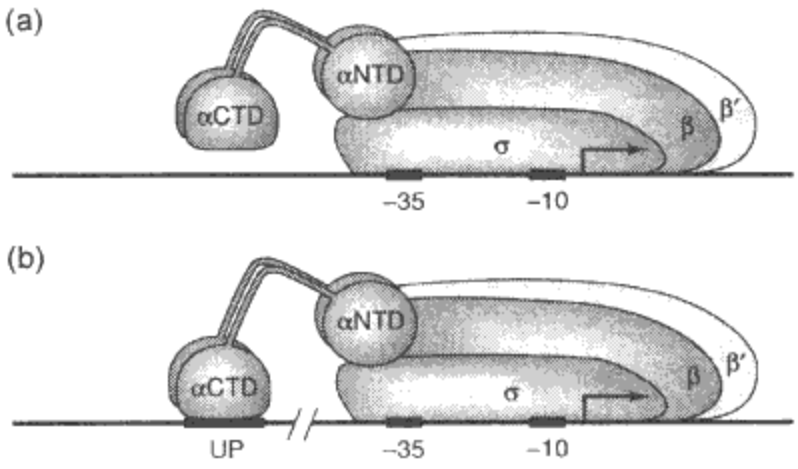


图 6.26　聚合酶 α 亚基 C 端结构域（CTD）的功能模型。（a）在核心启动子中，α 亚基不参与结合；（b）在含有 UP 元件的启动子中，CTD 与 UP 元件结合。注意，2 个 α 亚基均在图中显示，另一个重叠在后面。

小结　RNA 聚合酶的 α 亚基拥有独立折叠的 C 端结构域，该结构域识别启动子的 UP 元件并与之结合，从而实现聚合酶与启动子之间的紧密结合。

6.4　延伸

转录起始完成后，核心酶继续延伸 RNA 链，将核苷酸逐个添加到正在生长的 RNA 链上。本节将探讨转录的延伸过程。

核心酶在延伸过程中的作用

到目前为止，我们一直关注 σ 因子，因为它在决定转录起始的特异性方面发挥了关键作用。然而，核心酶拥有 RNA 合成装置，是延伸的核心执行者。在本节我们将了解 β 亚基和 β′亚基参与磷酸二酯键形成、参与聚合酶与 DNA 结合的证据，以及 α 亚基的多种活性（包括核心酶组装）。

β 亚基在磷酸二酯键形成过程中的作用　1970 年，Walter Zillig 首次对核心酶的单个亚基进行研

究。先从 *E. coli* 核心酶中分离出 3 个多肽组分，再重新组装成有活性的聚合酶。分离过程如下：Alfred Heil 和 Zillig 在含尿素的醋酸纤维素中电泳核心酶。与 SDS 类似，尿素也是一种变性剂，能够从蛋白质复合体中分离出单个多肽，但与 SDS 不同，尿素是一种温和变性剂，相对容易从蛋白质中除去。因此，经尿素变性的多肽比经 SDS 变性的多肽容易复性。电泳结束后，Heil 和 Zillig 切下含有聚合酶亚基的醋酸纤维素条，通过离心，使纤维素上的缓冲液和蛋白质析出。研究者共获得 3 个可分离的多肽，电泳检测其纯度（图 6.27）。

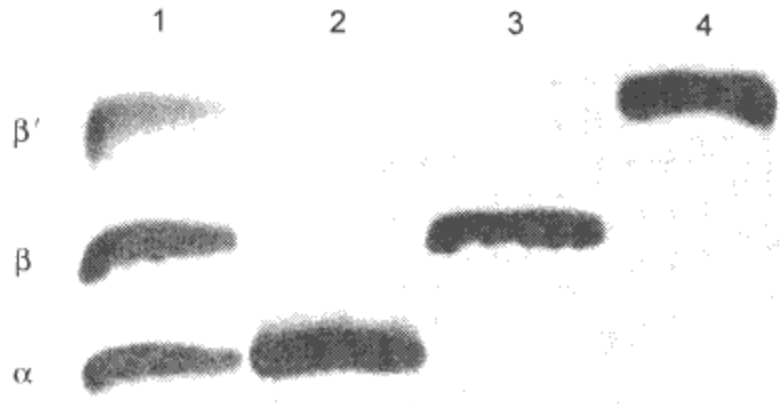


图 6.27　*E. coli* RNA 聚合酶单个亚基的纯化。Heil 和 Zillig 将 *E. coli* 的核心酶经尿素变性后，在醋酸纤维素膜上进行凝胶电泳，然后收集分离的多肽。泳道 1：电泳后的核心酶；泳道 2：纯化的 α 亚基；泳道 3：纯化的 β 亚基；泳道 4：纯化的 β′亚基。（*Source*：Heil，A. and Zillig，W. Reconstitution of bacterial DNA-dependent RNA-polymerase from isolated subunits as a tool for the elucidation of the role of the subunits in transcription. *FEBS Letters* 11（Dec 1970）p. 166，f. 1.）

一旦分离获得亚基，研究者将这些亚基重新组装成有活性的聚合酶，在 σ 因子存在条件下，组装过程最为有效。利用分离—重组系统，Heil 和 Zillig 混合并匹配不同来源的组分，以此获知各亚基的功能。例如，已知核心酶决定 RNA 聚合酶对抗生素利福平的敏感性或抗性，且利福平可阻断转录的起始。核心酶分离-重组实验的目的是明确到底是哪一个亚基赋予聚合酶对抗生素的敏感性或抗性。当 Heil 和 Zillig 把来自利福平敏感细菌的 α、β′和 σ 亚基与来自利福平抗性细菌的 β 亚基组合时，形成的聚合酶对利福平具有抗性（图 6.28）；与之相反，只要 β 亚基源于利福平敏感菌株，不管其他亚基来源如何，重组酶都是利福平敏感的。因此，显然是 β 亚基决定着聚合酶对利福平的敏感性或抗性。

另一种众所周知的抗生素链霉素（streptolydigin）能够阻断 RNA 链的延伸。用同样的分离-重组策略，Heil 和 Zillig 研究表明，β 亚基同样也决定聚合酶对链霉素的敏感性或抗性。初看起来似乎这一结果是矛盾的，同一个核心酶亚基是如何既参与转录的起始又参与转录的延伸？答案在于，利福平阻断了早期的延伸，使 RNA 链的长度不超过 2 或 3 个核苷酸，我们将在本章后面内容中详尽讨论这个问题。严格来讲，利福平是阻止转录的起始，因为在 RNA 长度超过 10 个核苷酸时，转录起始才算结束。然而，利福平确实影响隶属起始范畴的延伸。

1987 年，M. A. Grachev 及同事用**亲和标记**（affinity labeling）技术为 β 亚基的假定功能提供了直接证据。该技术的思路是用常规底物的衍生物标记酶，该衍生物可与蛋白质交联，用亲和试剂找出并标记酶的活性位点，最后，解离酶，观察标签结合的亚基。Grachev 及同事选用的 14 种不同亲和试剂均为 ATP 或 GTP 的类似物（analog），其中第一个类似物称为 I，其结构如图 6.29a 所示。当类似物 I 与 RNA 聚合酶混合后，会定位到酶的活性位点，如同 ATP 起始转录的行为，在活性位点与氨基作用形成共价键，该反应如图 6.29b 所示。

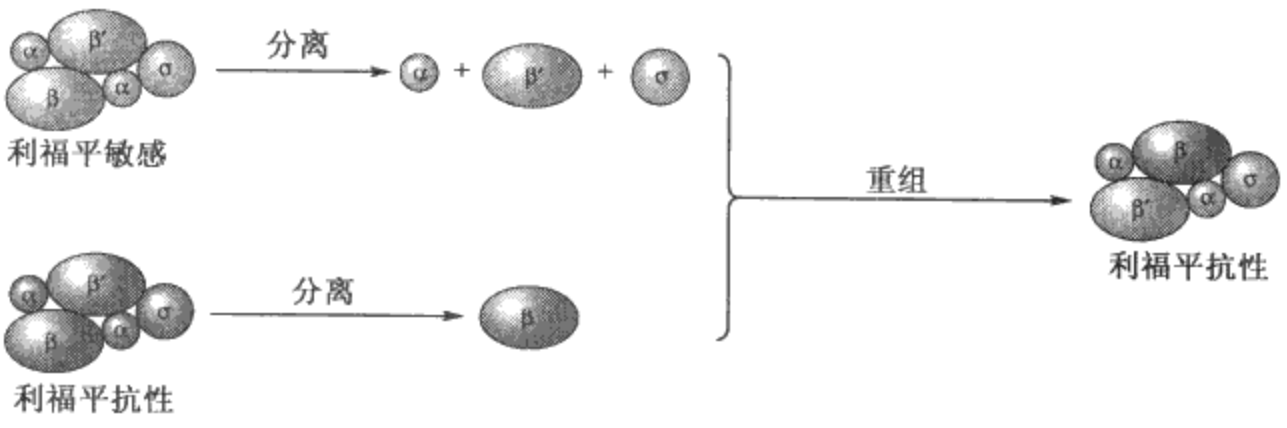


图 6.28　RNA 聚合酶的分离重组实验定位决定抗生素抗性的亚基。分别从对利福平敏感和抗性 *E. coli* 细胞中提取 RNA 聚合酶，分离多肽组分，并以多种组合方式重新组装成有活性的酶。实验中，当 α、β′以及 σ 亚基来源于对利福平敏感的聚合酶（蓝色），而 β 亚基来源于对利福平抗性的聚合酶（红色）时，重组酶具有利福平抗性，表明 β 亚基决定聚合酶对利福平的敏感性或抗性。

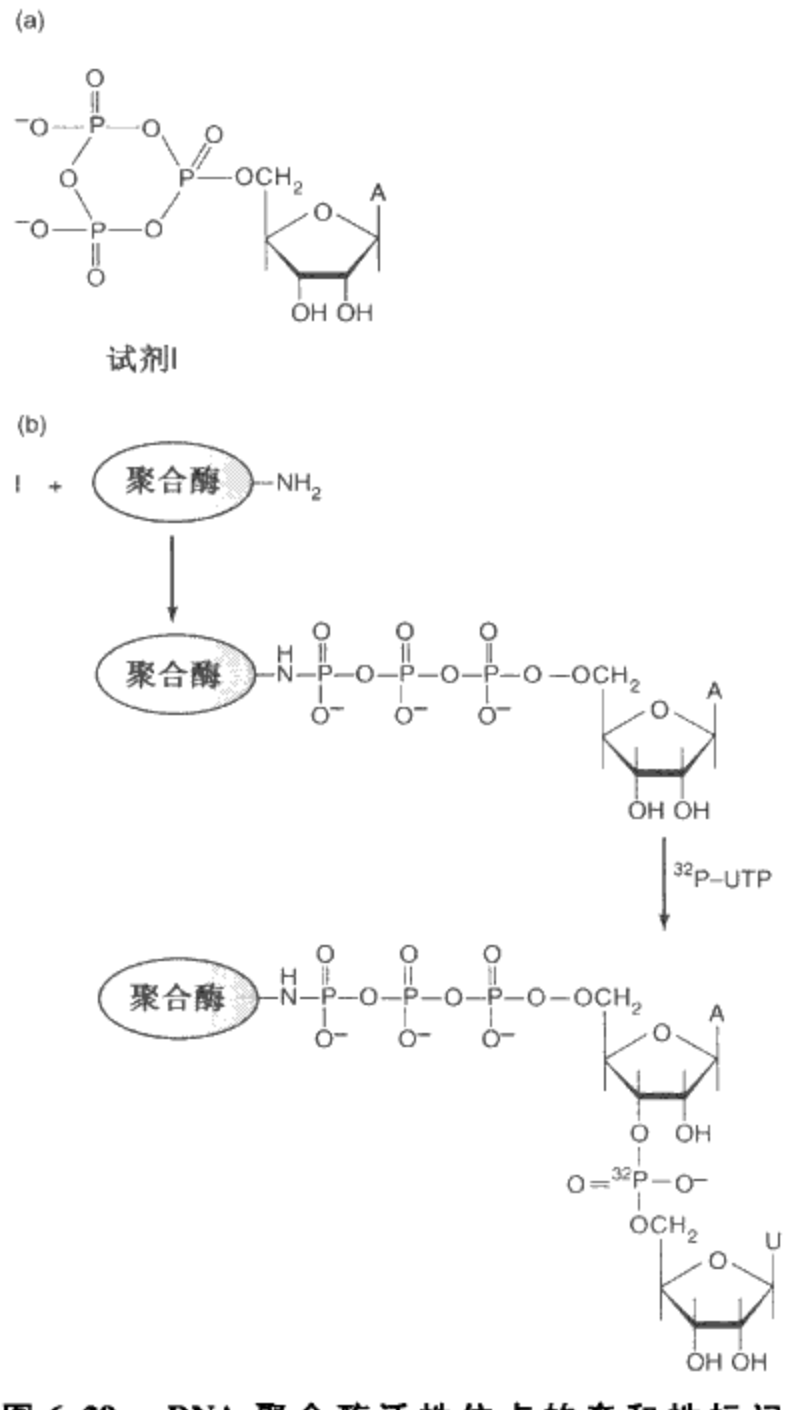


图 6.29　RNA 聚合酶活性位点的亲和性标记。（a）亲和试剂 ATP 类似物 I 的结构。（b）亲和标记反应。将亲和试剂 I 与 RNA 聚合酶混合，亲和试剂就会与活性位点的氨基发生共价结合（可能还有其他位点），然后加入放射性标记的 UTP，它与结合在酶上的亲和试剂 I 化合形成磷酸二酯键（蓝色）。该反应仅发生在活性位点，从而放射性标记活性位点。

理论上讲，研究者可以标记亲和试剂，并由此展开研究。然而，这一简单策略存在缺陷，即除了活性位点之外，亲和试剂还可能与酶表面的其他氨基结合。为解决这一问题，研究者先用非标记的亲和试剂与聚合酶结合，然后用放射性标记的核苷酸（[α-^{32}P] UTP 或 CTP）与结合在活性位点的亲和试剂形成磷酸二酯键，从而使活性位点被标记，而酶的其他位点不会被标记。最后解离被标记的酶，并对其亚基进行 SDS-PAGE 分析，图 6.30 显示了实验结果。很显然 β 亚基是唯一被亲和试剂标记的核心酶亚基，表明该亚基位于或十分邻近磷酸二酯键形成的位点，在有些实验中，可观察到标记 σ 因子，说明该亚基可能也邻近催化中心。

小结　核心酶 β 亚基位于 RNA 聚合酶催化磷酸二酯键形成的活性位点附近，σ 因子也位于核苷酸结合位点附近，至少在转录起始阶段情况如此。

β 亚基和 β′亚基在 RNA 聚合酶与 DNA 结合中的作用　1996 年，Evgeny Nudler 及同事通过一系列巧妙实验证实 β′和 β 亚基均参与了 RNA 聚合酶与 DNA 的结合。在 DNA 分子上有两个结合位点：①上游弱结合位点，是 DNA 解链区，主要通过静电作用与 RNA 聚合酶结合；②下游强结合位点，通过疏水作用使酶与 DNA 结合。

图 6.31 示意了 Nudler 及同事在实验中所采用的常规方法。用第 4 章提到的基因克隆技术，研究者使 RNA 聚合酶 β′亚基 C 端连接 6 个额外的组氨酸（histidine），将此聚合酶吸附在镍树脂上，通过洗涤树脂和加入新鲜试剂，可快速更换模板和底物。图 6.31a 显示如何通过添加某些核苷酸（如 ATP、CTP 和 GTP，但不是 UTP）而使聚合酶“步移”到模板的特定位点处，从而合成短的 RNA 产物，并因停留在需要缺乏核苷酸的位点处。洗除第一组核苷酸，加入第二组核苷酸，使聚合酶沿模板步移到下游更远的某一位点处，如此反复。最后，研究者加入所有的 4 种核苷酸，将聚合酶从模板上“驱赶”下来，并合成截断转录物。

图 6.31b 显示了一个惊人现象：当 Nudler 及同事在第二模板及 ATP、GTP、CTP 存在条件下驱赶聚合酶时，聚合酶可从原初模板（蓝色）跳转到第二模板（紫红色）上，继续延伸转录物直至到达模板链的 A 位点处。此时，聚合酶因 UTP 的缺乏而不得不停止延伸过程。研究者之所以知道聚合酶可从一条模板链跳转到另一条模板链，是因为可将聚合酶从第二模板链上驱赶下来，并得到预期大小的截断转录物，且比原初模板的截断转录物要长。更令人信服的是，研究者使聚合酶沿第二模板步移时，获得了预期大小的捕获转录物（arrested transcript）。

其次，研究者改变第二模板的特征，以便了解聚合酶向第二模板跳转并继续转录的必需条件。他们发现下列特征。①单链的第二模板只有在低盐浓度下才能作为转录的模板，在高盐浓度下复合体解离，表明聚合酶不能紧密结合单链 DNA。②抗盐性聚合酶与 DNA 结合时只要求 9bp 的双链 DNA，且该双链 DNA 必须在相对于转录物生长的 3′端的+2 和+11 位点之间。这种抗盐性结合表明是疏水性作用而不是静电作用的结果，即这种结合涉及聚合酶的疏水氨基酸与 DNA 大沟内碱基的相互作用。③盐稳

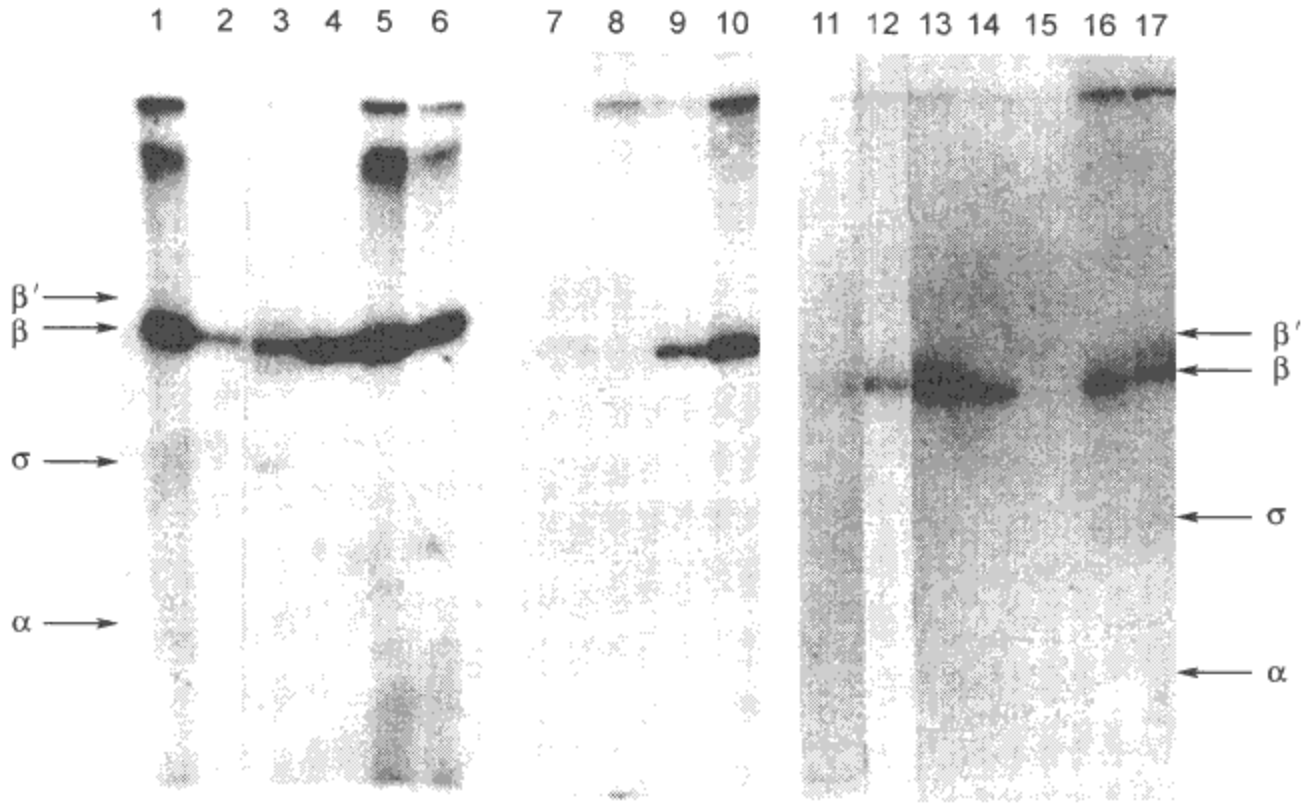


图 6.30　β亚基位于或邻近形成磷酸二酯键的活性位点。Grachev 及同事用图 6.29 所示的方法标记 *E. coli* RNA 聚合酶的活性位点，通过电泳分离聚合酶亚基，以确定活性位点的亚基组成。除泳道 5、6 外，各泳道表示用不同核苷酸亲和试剂和放射性 UTP 的实验结果。泳道 5、6 用相同的亲和试剂，只是泳道 5 使用放射性 UTP，泳道 6 使用放射性 CTP，其他泳道均使用放射性 UTP。放射自显影结果显示 β 亚基被大多数亲和试剂标记，但有时 σ 亚基也被微弱标记。因此，β 亚基位于或邻近催化磷酸二酯键形成的活性位点处。（*Source*：Grachev et al.，Studies on the functional topography of *Escherichia coli* RNA polymerase. *European Journal of Biochemistry* 163（16 Dec 1987）p. 117，f. 2.）

定性结合的另一个必要条件是在 RNA 的 3′端上游存在一段长度至少为 6nt 的单链 DNA。

再次，研究者用聚合酶已步移到接近下游末端的原初模板链分析了盐敏感性结合位点。这些复合体缺少盐抗性结合所必需的下游双链区，因此保留了全部的盐敏感性结合位点。盐敏感性结合的必需条件是什么呢？基本要求是在相对于 RNA 3′端 −6～+1 区间内存在至少为 6nt 的单链 DNA。

图 6.32 总结了这些研究发现。下游的盐稳定性结合位点（黄色）在+2～+11，上游的盐敏感性结合位点（蓝色）位于−6～+1。盐敏感性结合位点由单链 DNA 组成，该单链 DNA 为模板链。这反映了这样一个事实：在包含 RNA 聚合酶的转录活性复合体（transcription−competent）中，该区域已被解链。该模型认为，RNA 转录物与 DNA 的相互作用并不是稳定 DNA-RNA-蛋白质三元复合体的最重要贡献者，因为这种相互作用与盐敏感（弱）结合位点有关。三元复合体稳定性的最主要贡献者是下游的强盐抗性结合位点。事实上，这个强结合位点是将聚合酶牢牢结合在 DNA 上的滑动钳（sliding clamp）关键部分。

通过研究，Nudler 及同事也确定了这两个结合位点所涉及的聚合酶亚基。其研究策略是：将胸腺嘧啶类似物 5-碘代脱氧尿嘧啶核苷酸（5-iododeoxyuridine）引入到较短且具有放射性标记的第二模板的特定位置处，经紫外线激活后，该核苷酸可与邻近的蛋白质交联。选择的研究位点是−3、+1 和+6 位，分别位于盐敏感区、活性位点和盐稳定区。根据定义，+1 位在活性位点，因为该位点对应于 RNA 的 3′端，是最后一个核苷酸刚被添加的部位。

交联后，研究者对复合体进行 SDS-PAGE 和放射自显影分析，确定与放射性标记 DNA 结合的聚合酶亚基。正如我们所预想的那样，图 6.33 显示 β 亚基在−3 位和+1 位与 DNA 交联，因为 β 亚基位于或邻近酶的活性位点。另一方面，β 和 β′亚基均与 DNA 在−6 位交联，因为这里是强结合位点，该结果与早期的发现相一致，即 β′亚基自身可与 DNA 紧密结合。图中还显示，β 亚基既能在紧密结合过程中发挥作用，也在活性位点的弱结合中起作用。此外，当单链 DNA 在+6 位与聚合酶交联时，很少发生与 β′亚基结合的情况，这也进一步证实通过 β′亚基介导的紧密结合需要双链 DNA。

最后，Nudler 及同事确定了位于 β 和 β′ 亚基内的 DNA 结合位点。研究者通过 SDS-PAGE 纯化与标记 DNA 交联的 β 和 β′亚基，然后用 CNBr 处理这些多肽，CNBr 能够在蛋氨酸（methionine，Met）残基之后切割多肽。在所产生的系列肽段中，有些与标记模板交联。凝胶电泳后，放射自显影的结果

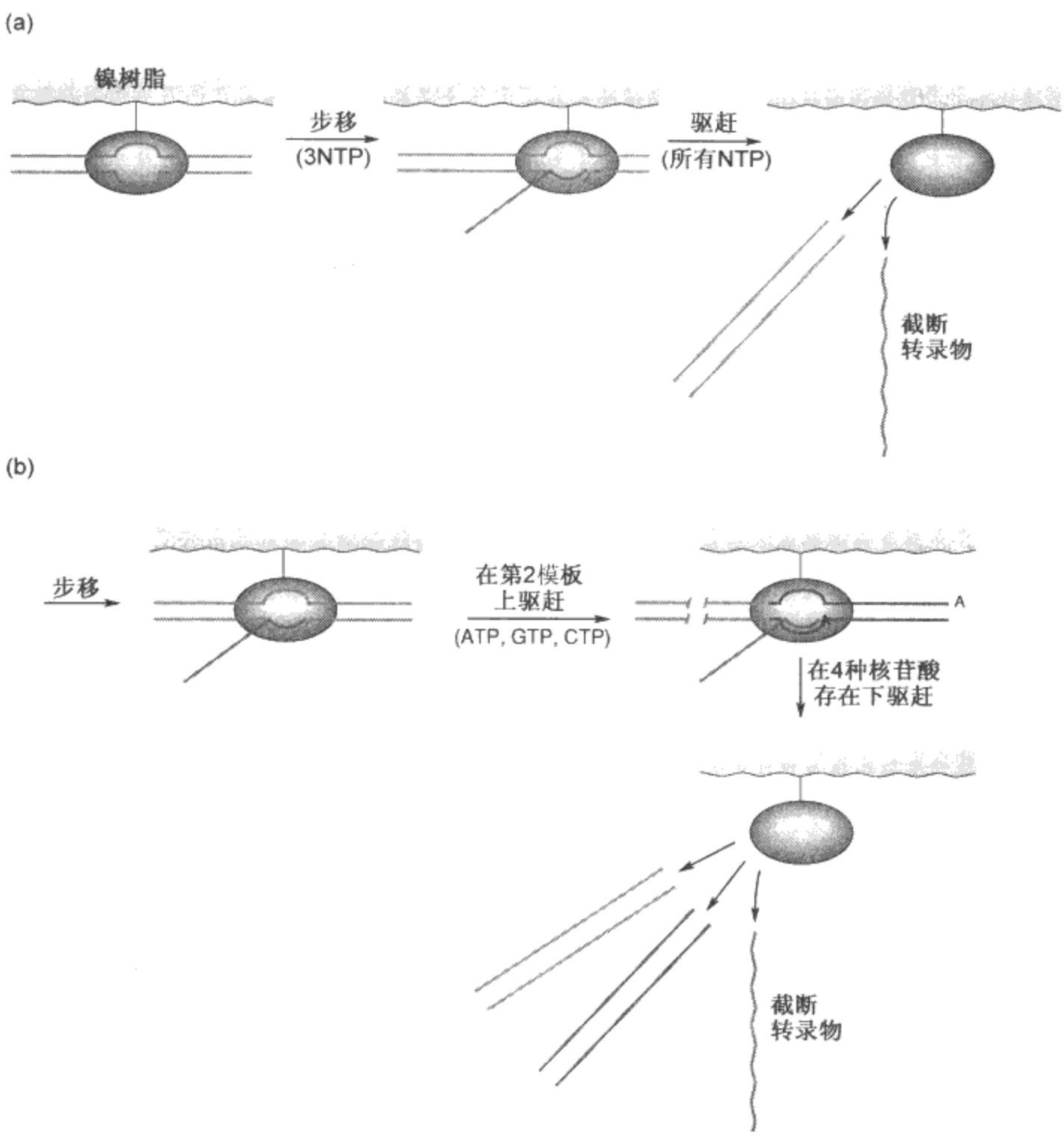
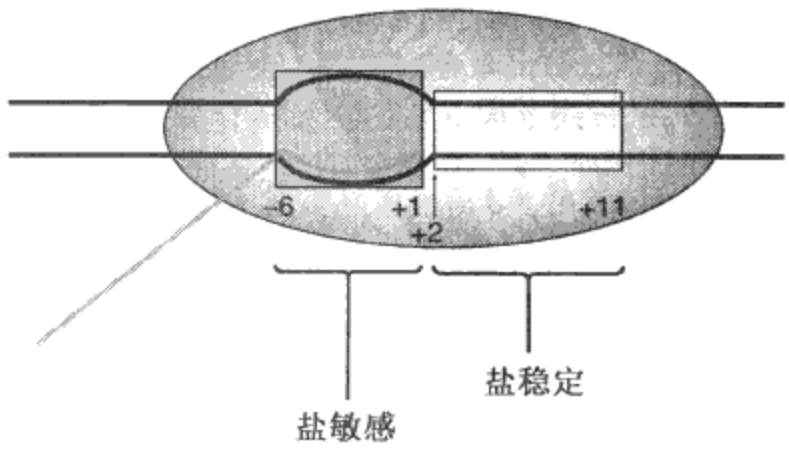


图 6.31　确定与 RNA 聚合酶结合所需模板条件的策略。（a）RNA 聚合酶的步移与驱赶（walking and chasing）。Nudler 及同事用寡聚组氨酸标签标记聚合酶的 β′亚基，使 RNA 聚合酶（红色）结合在镍树脂（黄色）上。加入几组 3 种或少于 3 种核苷酸，使 RNA 聚合酶沿 DNA 模板向前步移几个碱基对，然后因缺乏某种核苷酸而停顿，最后，加入所有的 4 种核苷酸，使 RNA 聚合酶一直步移到 DNA 模板的末端并脱离下来，释放截断转录物。（b）模板转换。Nudler 及同事再次使聚合酶步移到原初模板（蓝色）的特定位点，然后，在第二个模板（紫红色）和 ATP、GTP、CTP 存在条件下驱赶聚合酶，使 RNA 聚合酶由原初模板跳转到第二个模板，并继续前行，直到需要聚合 UTP 的位点（模板链上的第一个 A）后才会停止。值得注意的是，聚合酶向互补链的跳转可被阻断，因为，如果上链是模板链，其第一个位置恰好需要缺乏的 UTP。最后，RNA 聚合酶被驱赶下第二模板链，释放出两个模板链和延长的截断转录物。

图 6.32　*E. coli* RNA 聚合酶全酶的 DNA 结合位点。基于图 6.31 所示的模板转换实验，只是用许多可改变的原初模板和第二模板。Nudler 及同事指出，在 DNA 模板上，有两个能够与 RNA 聚合酶发生相互作用的位点：下游位点（黄色），为＋2 位至＋11 位之间的序列，两者靠疏水作用而结合，在盐溶液中是稳定的；上游位点（蓝色），位于熔解 DNA 区间内，从－6 位至＋1 位，两者靠静电作用而结合，在盐溶液中是不稳定的。



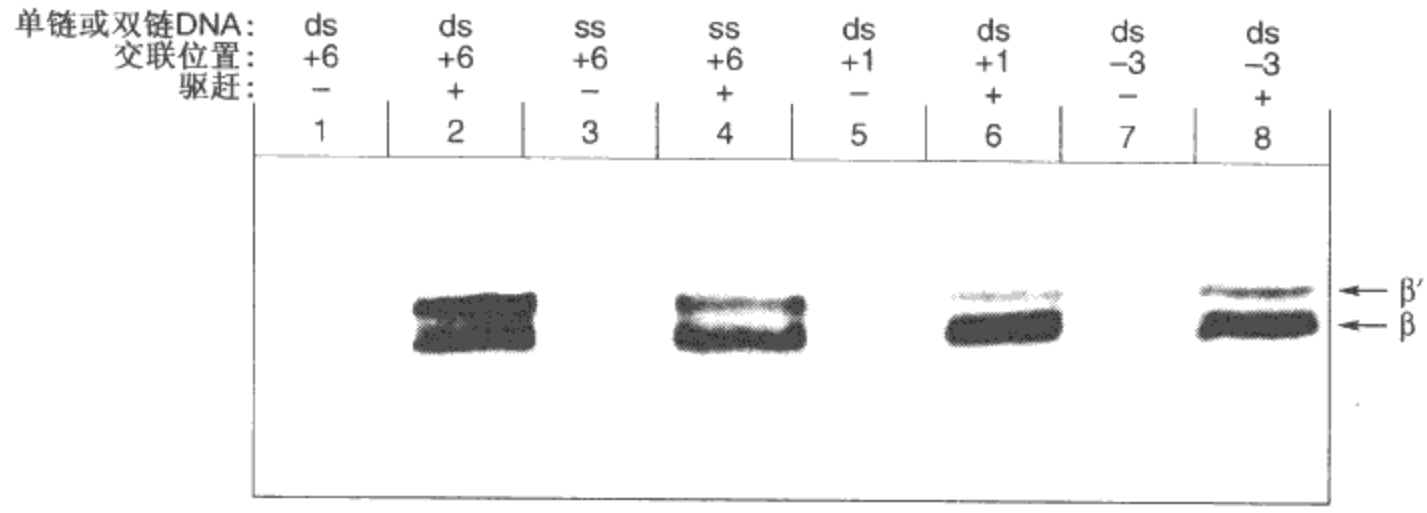


图 6.33　标记 DNA 与 RNA 聚合酶亚基的交联。Nudler 及同事将反应活性核苷酸（5-碘代脱氧尿嘧啶核苷酸）在交联位点处引入单链（ss）或双链（ds）标记的第二模板中，交联位点及第二模板在图上方标注。如图上方所示，研究者或者将聚合酶从原初模板驱赶（＋）到第二模板上，或不进行驱赶（－）。驱赶（不驱赶）后，用紫外线照射聚合酶-DNA 复合体，使反应活性核苷酸与邻近的聚合酶亚基交联。对交联产物进行 SDS-PAGE 和放射自显影处理，显示标记的聚合酶亚基。只有被驱赶到第二模板上的聚合酶才会发生交联，β 亚基参与了酶与所有 DNA 位点的结合，而 β′亚基主要参与酶与双链 DNA 在＋6 位点处的结合，当单链 DNA 在＋6 位与聚合酶交联时，很少发生与 β′亚基的结合。（*Source*: Nudler et al., Transcription processivity: Protein-DNA interactions holding together the elongation complex. *Science* 273 (12 July 1996) f. 5c, p. 214. © AAAS.）

显示，与 β 亚基的交联发生在 N 端附近，位于 Met30 和 Met102 之间，而与 β′亚基的交联发生在 C 端附近，位于 Met 1230 和 Met1273 之间。

小结　模板转换实验明确表明，与聚合酶发生相互作用的 DNA 位点有 2 个。一个为弱结合位点，包括 DNA 的解链区和位于或靠近聚合酶 β 亚基的催化位点，在此位点发生的蛋白质-DNA 之间的相互作用主要依赖于静电作用，具有盐敏感性；另一个是强结合位点，涉及活性位点的下游 DNA 和聚合酶的 β 亚基和 β′亚基。

延伸复合体的结构

在本节，我们将学习掌握如何通过结构研究而完美地预测核心聚合酶亚基的功能。同时还将学习转录延伸的拓扑学特征，即聚合酶如何解决模板链的解旋与再旋问题，以及如何沿着缠绕（螺旋）的模板移动而不会使 RNA 产物缠绕在模板链上。

RNA-DNA 杂交　关于这一点，我们假设 RNA 产物在脱落、排出聚合酶之前，会通过几个碱基与其 DNA 模板链形成 RNA-DNA 杂合分子。然而有关杂合分子的长度一直存在争议，估计其长度为3～12bp，有些研究者甚至怀疑杂合分子的存在。Nudler和 Goldfarb 及同事应用转录物步移技术和 RNA-DNA 交联技术，证明在延伸复合体中确实存在着 RNA-DNA 杂合分子，其长度为 8bp 或 9bp。

研究者将在 UMP 的类似物（U·）引入^{32}P 标记的新生 RNA 产物 5′端的第 21 位或 45 位。正常情况下 U·无反应活性，但在 $NaBH_4$ 存在时 U·可与配对的碱基交联，如图 6.34a 所示。实际上，U·能够接触到 DNA 链上与配对碱基 A 相邻的一个嘌呤碱基，但本实验的设计方案可以阻止这种情况的发生。所以，交联只能发生在 DNA 模板链上的 A 与转录物 RNA 的 U·之间，如果没有碱基配对，就不可能发生交联。

Nudler、Goldfarb 及同事将转录物中的 U·步移到距 3′端的不同位置，从－2 位（紧邻 3′端的核苷酸，3′端核苷酸标记为－1）开始延伸到－44 位。然后，研究者使 RNA 与 DNA 模板链交联，在同一块凝胶上电泳 DNA 和蛋白质，而在另一块凝胶上电泳 RNA。结果表明，RNA 总能被标记，而 DNA 或蛋白质只有与 RNA 发生后才能被标记。

图 6.34b 显示了实验结果。如果碱基 U·出现在－2～－8 位，则 DNA 会被强烈标记，若 U·出现在－10 位或更远处则 DNA 只能会被微弱标记。因此，碱基 U·只有存在于－2～－8 位时，才能与 DNA 模板链上的 A 互补配对，当 U·在－10 位时，碱基配对能力会被严重削弱。所以，RNA-DNA 杂交链从－1 位延伸到－8 位或－9 位，不会在向外延伸了（在 RNA3′端－1 位的碱基必须与模板链上碱基准确配对，以便在 RNA 合成时能够引入正确的碱基）。这一结论通过蛋白质标记实验得到进一步验证。当 U·不在杂交区域（－1～－8 位）时，RNA

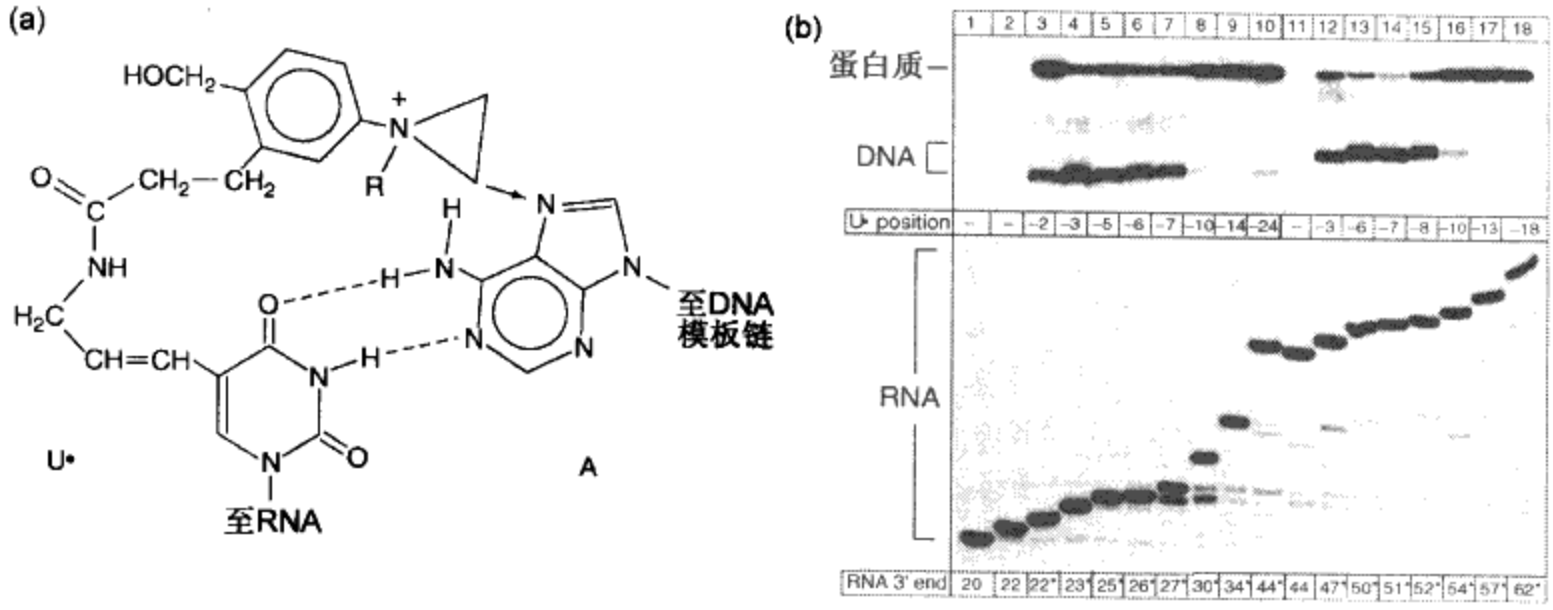


图 6.34　在延伸复合体中 RNA-DNA 及 RNA-蛋白质间的交联。（a）交联试剂 U·与 DNA 模板链上的 A 发生碱基配对后的结构。箭头标注了该交联试剂与 DNA 形成的共价键。（b）交联结果。Nudler 和 Goldfarb 及同事将延伸复合体中^{32}P 标记的新生 RNA 产物的第 21 位或 45 位引入 U·，然后让 U·移动到新生 RNA 3′端（−1 位）的−2～−24 不同的位点，然后使 RNA 与 DNA 模板（或 RNA 聚合酶中的蛋白质）交联。在同一凝胶上电泳交联的 DNA 和蛋白质，在另一凝胶上电泳游离的 RNA。放射自显影检测。泳道 1、2 和 11 为 RNA 中无 U·的阴性对照；泳道 3～10：在初生 RNA 产物的 21 位有 U·；泳道 12～18：在初生 RNA 产物的 45 位有 U·。底部的星号表示 RNA 中存在 U·，只有当 U·出现在−2～−8 位的位点时，才会发生 RNA 与 DNA 间的交联。（*Sources*：(a) Reprinted from *Cell* 89. Nudler, E. et al. The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase fig. 1, p. 34 © 1997 from Elsevier (b) Nudler, E. et al. The RNA-DNA tybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell* 89 (1997) f. 1, p. 34. Reprinted by permission of Elsevier Science.）

聚合酶中的蛋白质就会被强烈标记。这可能反映了这样的事实，即活性基团在不能与 DNA 模板链进行碱基互补配对时，更易与蛋白质结合。最近对 T7 RNA 聚合酶的研究结果显示，RNA-DNA 杂合分子的长度为 8bp。

小结　在 *E. coli* 延伸复合体中，相对于初生 RNA 产物的 3′端，RNA-DNA 杂合分子的长度从−1 位延伸到−8 位或−9 位。T7 的杂合分子的长度为 8bp。

核心聚合酶的结构　为了得到最清晰的延伸复合体结构图像，我们需要知道核心聚合酶的结构。X 射线晶体学图像能够提供最高的分辨率，但需要获得三维结构的晶体，而到目前为止尚未成功获得 *E. coli* 聚合酶的三维晶体。然而，1999 年 Seth Darst 及同事成功结晶出嗜热水生菌（*Thermus aquaticus*）的核心聚合酶，晶体结构的分辨率达 3.3Å。该结构整体上十分类似于电子显微镜观察到的 *E. coli* 核心酶低分辨率二维晶体结构，这意味着两种酶在细微结构上也可能十分相似。换言之，*Thermus aquaticus* 的聚合酶晶体结构是我们目前了解细菌核心酶结构的最好窗口。

图 6.35 从 3 个不同方位描述了酶的整体轮廓。首先值得注意的是，*Thermus aquaticus* 的 RNA 核心酶形似一只张开的蟹钳，为了便于区分，用不同的颜色表示核心酶的 4 个亚基（β、β′和 2 个 α 亚基）。彩图显示，蟹钳的一半主要由 β 亚基组成，另一半主要由 β′亚基组成，2 个 α 亚基位于蟹钳的节点处，其中之一（αI，黄色）与 β 亚基相连，另一个 α 亚基（αII，绿色）与 β′亚基相连，较小的 ω 亚基位于底部，覆盖于 β′ 亚基的 C 端。

图 6.36 显示了核心聚合酶的催化中心（catalytic center）。在核心酶的两钳之间形成一宽约 27Å 的通道，DNA 模板可能位于这个通道内。酶的催化中心用 Mg^{2+} 标记（粉红色球体）。有三方面的证据表明 Mg^{2+} 位于酶的催化中心。第一，到目前为止，所有被研究的细菌 β′ 亚基都有一段保守的氨基酸序列 NADFDGD，推测其中的 3 个天冬氨酸（D）的作用可能是螯合 Mg^{2+}。第二，3 个天冬氨酸残基中的任何一个发生突变都是致死的，它们使酶在启动子处形成开放启动子复合体，但无催化活性。因此，这 3 个天冬氨酸残基是催化活性而不是紧密结合 DNA 所必需的。最后，如图 6.36 所示，*T. aquaticus*

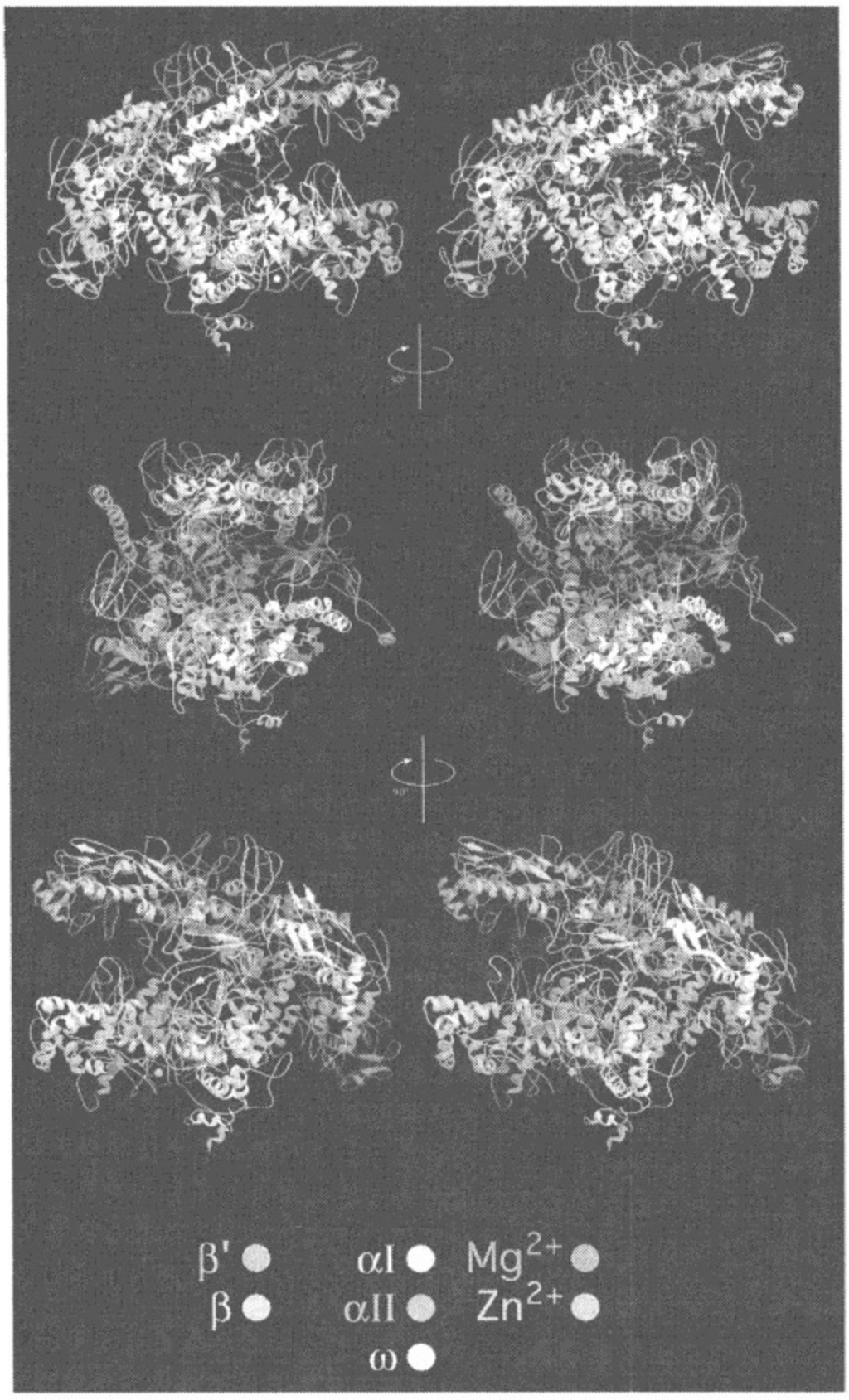


图 6.35　*T. aquaticus* RNA 聚合酶核心酶的晶体结构。图中显示了 3 种不同的立体图，区别在于相差 90°的旋转。聚合酶亚基和金属离子用不同的颜色标注，如图底部所示。金属离子表示为彩色小球，较大的红色和绿色圆点分别表示缺失的 β、β′亚基的未结构折叠区。（*Source*: Zhang, G. et al., Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3Å resolution. *Cell* 98 (1999) 911-24. Reprinted by permission of Elsevier Science.）

的核心聚合酶晶体结构表明，3 个天冬氨酸残基（红色）侧链确实与 Mg^{2+} 配位。所以，酶的催化中心有 3 个天冬氨酸残基和 Mg^{2+}。

图 6.36 也显示，在形成贯穿酶通道顶部的 β 亚基中存在利福平结合位点。发生改变可引起对利福平抗性变化的氨基酸用粉红色圆点标注，很明显，在三维结构中这些氨基酸紧密串联在一起，可能位于利福平结合位点。我们也知道，利福平允许 RNA 合成的起始，但在加入几个核苷酸后就会阻断 RNA 链的延伸。此外，一旦完成启动子清除，抗生素利福平就不会再对转录延伸产生任何影响。

如何依据抗生素的活性解释利福平结合位点的定位？一种假设认为，结合在通道内的利福平关闭了正在生长的 RNA 链的出口通道，阻断了短链 RNA 的生长。一旦 RNA 延伸到特定的长度，就会阻断利福平与利福平结合位点的靠近，或至少能够

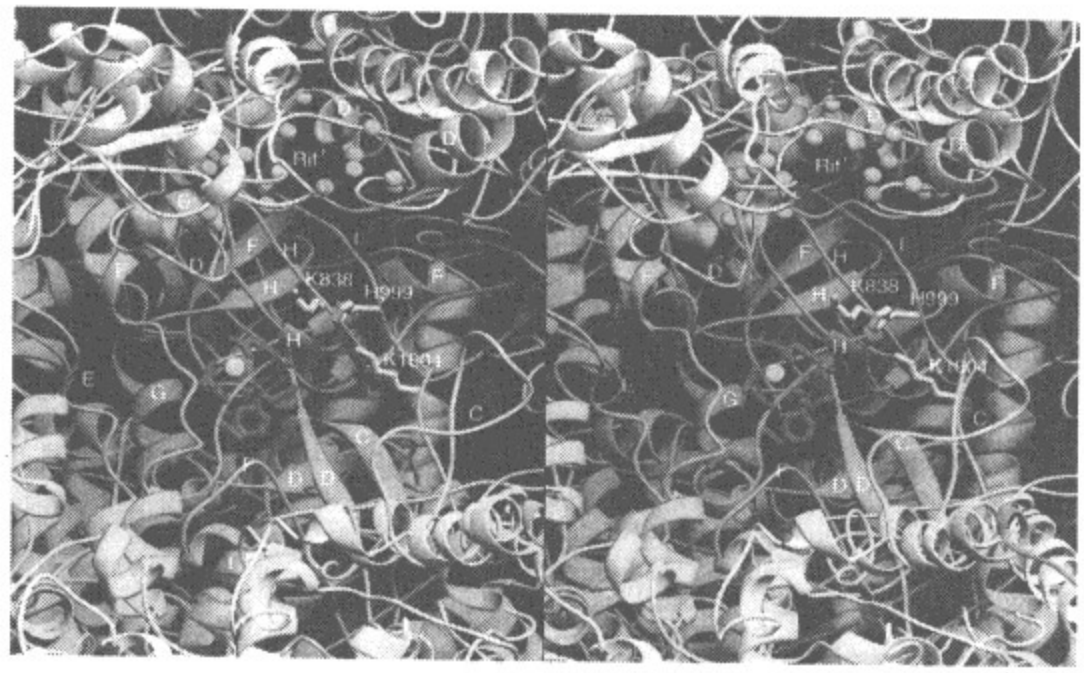


图 6.36　核心聚合酶的催化活性中心。粉红色球体代表 Mg^{2+}，与 3 个天冬氨酸侧链（红色）配位，位于通道顶部、参与决定利福平抗性的氨基酸用紫色球体表示，这些氨基酸环绕着假定的利福平结合位点或 Rif 口袋，用 Rif' 表示。聚合酶亚基的颜色与图 6.35 中的相同（β'为粉红色；β 为青绿色；α 为黄色和绿色）。该图中的两张图版是立体图像的两半。（*Source*：Zhang，G. et al.，Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3Å resolution. *Cell* 98 (1999) 911-24. Reprinted by permission of Elsevier Science.）

有效阻止抗生素利福平的结合。

Darst 及同事通过 *T. aquaticus* 核心聚合酶与利福平复合体的晶体结构证实了这一假设。当 RNA 链延伸长度达 2nt 或 3nt 时，位于预知结合位点的利福平通过阻断延伸转录物的排出而阻断转录。

小结　*T. aquaticus* 的 RNA 核心聚合酶的 X 射线晶体研究发现，其形状类似一只蟹钳，能够抓牢 DNA 链。横贯核心酶的通道含有催化中心（Mg^{2+} 与 3 个天冬氨酸配位）和利福平结合位点。

全酶的结构　2002 年，Darst 及同事发表了 *T. aquaticus* RNA 聚合酶全酶分辨率为 4Å 的晶体结构图像。该研究得出三点结论。

1. σ 亚基（σ^A）与核心酶的 β 和 β'亚基间有较大的接触面。

2. σ 亚基缺失前 91 个氨基酸后，全酶结构中有一个过于狭窄的裂隙，以至于不能接纳 DNA 模板，这表明 σ 亚基缺失的结构域可能帮助撬开酶的裂隙以便与 DNA 结合。

3. σ 亚基中连接其两个功能域的中间无规则环明显靠近酶的活性位点，位于排出 RNA 产物的通道内。

后一个发现有两个含义。首先，由于十分靠近于活性位点，从而使靠近活性位点的 σ 因子部分在形成第一个磷酸二酯键时发挥作用，这与 σ 因子的转录起始功能相一致；其次，σ 因子对 RNA 排出通道的阻塞，是全酶在真正起始转录之前形成流产性转录物的原因。现在，让我们逐一分析这些结论的证据。

图 6.37a 显示了两个不同视角的全酶晶体结构。在左图中，我们可以看到酶的整体结构，核心亚基用不同的色彩显示，α 碳骨架代表 σ 亚基，其中的 α 螺旋用圆柱体表示。在右图中，我们可以看到部分 β'亚基被略去的酶结构，从而可看清 σ 亚基穿越酶的途径。在图 6.37b 的示意图中，σ 亚基的各部分用不同的色彩标注区分。σ 亚基与 β 和 β'亚基间的接触面分别用绿色和红色表示。

Darst 及同事未能获得完整全酶的结晶体，只能以结合缺失前 91 个氨基酸（结构域 1.1）的 σ 亚基的核心酶开展研究，缺失整个 σ 结构域后，我们虽不能了解全酶的形状，但还是可以做出一些预测。例如，图 6.38a 显示，短截了的 σ 亚基 N 端位于 α 螺旋区的末端，该 α 螺旋直接指向酶的活性中心。表明 $\sigma_{1.1}$ 结构域位于图 6.38b 左侧所示的位置。如果确实如此，就可以解释晶体结构的另一个特征：位于活性位点处的裂隙过于狭窄，以至于不能接纳 DNA 模板。综合这两方面的信息，Darst 及同事推测，包含 $\sigma_{1.1}$ 区在内的整个 σ 亚基可将酶的裂隙撬开使双链 DNA 模板结合在这个通道中，形成闭合启动子复合体，然后，在形成开放启动子复合体时，$\sigma_{1.1}$ 区被逐出，从而使主通道在活性位点处与解链 DNA 紧密结合。

这一见解得到了一系列证据的支持。首先，$\sigma_{1.1}$ 区为强酸性（在中性 pH 条件下，有 1/3 氨基酸残基带负电荷），有利于 $\sigma_{1.1}$ 区与包含碱性氨基酸残基

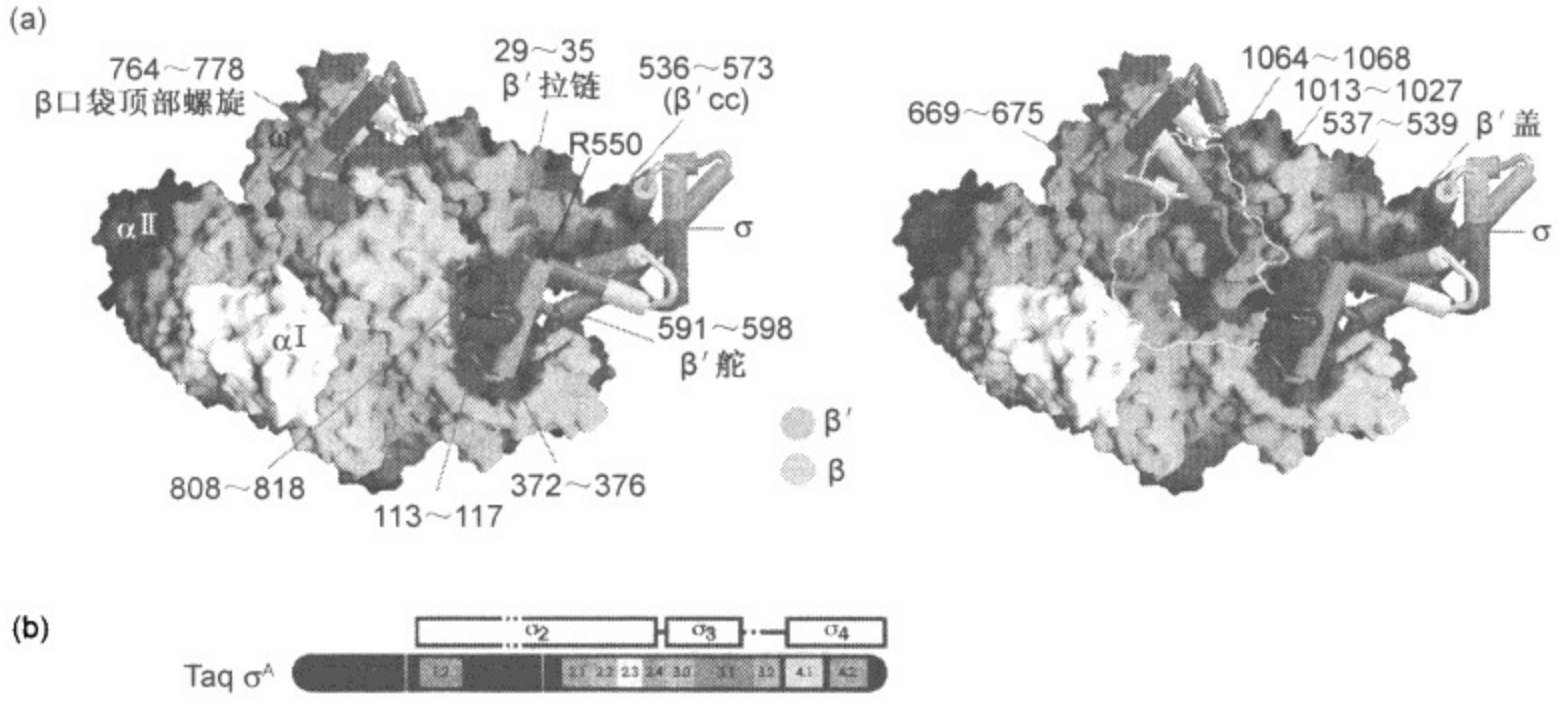


图 6.37　嗜热水生菌 RNA 聚合酶全酶的晶体结构。（a）聚合酶全酶结构的两种视图。ω、αI 和 αII 亚基位于灰色阴影区，β 和 β′亚基分别用蓝色和粉红色表示，σ 亚基用一组 α 螺旋（圆柱体）及紧随 α 碳骨架的环表示。不同颜色所代表的元件已在图（b）中的简图中标注。在图（a）的右侧结构图中，部分 β 亚基被移除，从而显示出 σ 亚基的方位，黄色线条勾勒出 β 亚基被移除的区域。与 σ 亚基相接触的 β 亚基和 β′亚基的区域分别用绿色和红色表示。（b）σ 亚基的保守区域用不同的颜色表示，这些区域被分成 3 个大的结构域：σ_2、σ_3 和 σ_4。（*Source*：Murakami et al.，*Science* 296：p. 1281. © 2002 by the AAAS.）

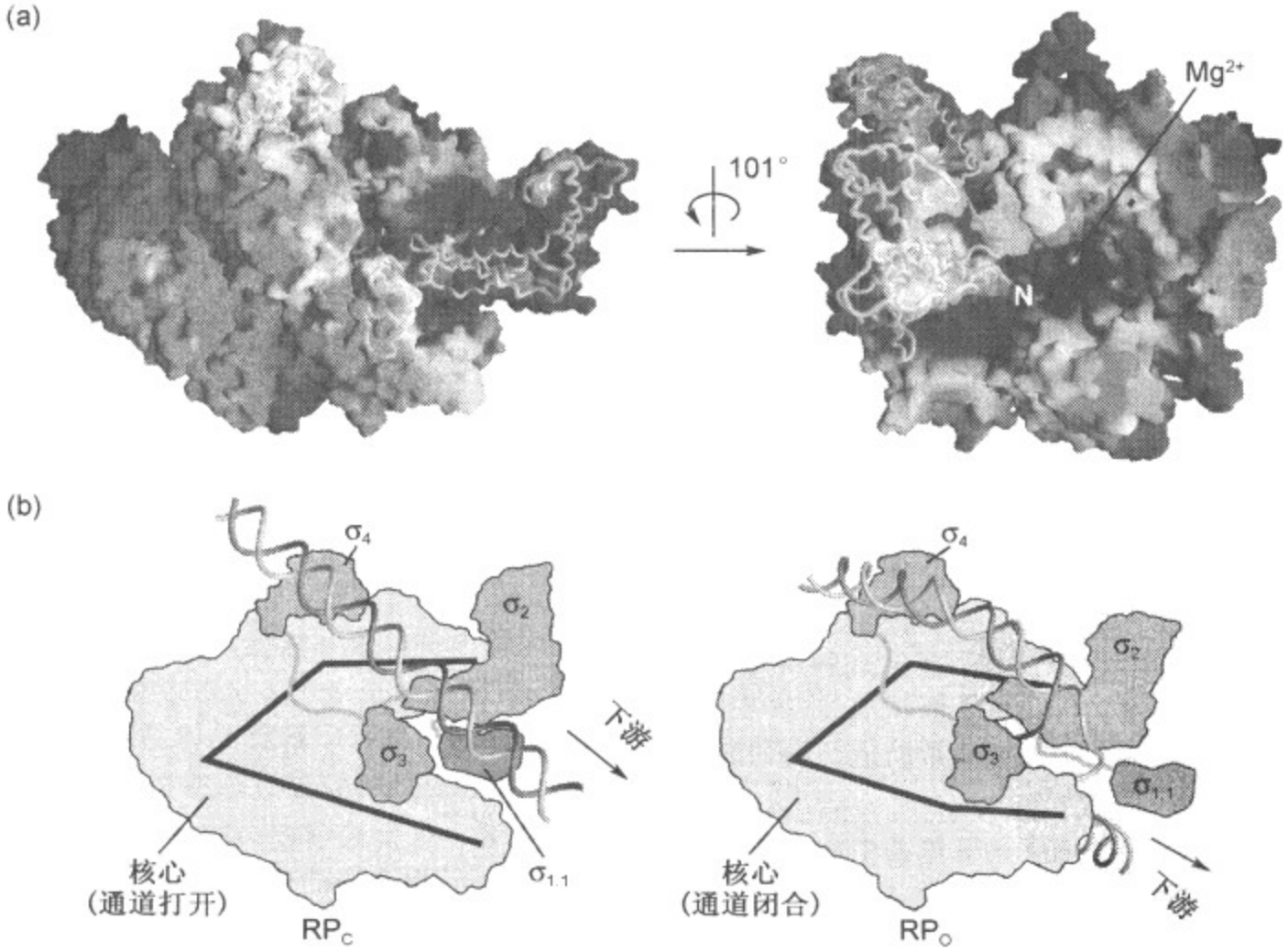


图 6.38　$\sigma_{1.1}$ 区域参与开放启动子复合体形成的模型。（a）全截短的全酶结构。左图，其观测方位同图 6.37a；右图，将左图旋转 101°，露出 σ 亚基的氨基末端（用“N”标记）。σ 亚基的 α 螺旋骨架（黄色）直接指向酶的主通道（蓝色）活性中心的 Mg^{2+}，这表明 $\sigma_{1.1}$ 区位于这一主通道内。（b）$\sigma_{1.1}$ 区的参与模型。在闭合的启动子复合体（RP_C）中，$\sigma_{1.1}$ 区位于酶的主通道内，使通道处于开放状态以接纳双链 DNA 模板（绿色）。在开放的启动子复合体（RP_O），$\sigma_{1.1}$ 区被逐出主通道，以便通道闭合包纳解旋的 DNA。（*Source*：Murakami et al.，*Science* 296：p. 1283. © 2002 by the AAAS.）

的主通道结合。其次，*E. coli* σ^{70} 的羟自由基蛋白质足迹表明，游离 σ 亚基的 $\sigma_{1.1}$ 区暴露在羟自由基中，但在全酶中是受保护的，一旦全酶形成开放启动子复合体，$\sigma_{1.1}$ 区就会暴露。再次，Darst 及同事获得了结合 DNA 的 *T. aquaticus* 全酶晶体。在该复合体中，1.1 区是无序的，这暗示它已被逐出主通道，正如模型所预测的。此外，晶体中的 1.1 区易被蛋白酶降解，再次表明 1.1 区在主通道中未受到保护。最后，Richard Ebright 及同事用 FRET 实验证实，在闭合启动子复合体中，1.1 区位于主通道内，但在开放启动子复合体中则被逐出主通道（如本章前面所述，1.1 区还有另外一个重要的作用行为，即无核心酶时阻止 σ 亚基与 DNA 结合）。

就转录起始而言，令人费解的情况之一是：为什么流产性转录是转录起始的显著特征？图 6.39 所描述的全酶结构可以给出解释，图 6.39a 的右侧着重显示全酶的 σ_3 区与 σ_4 区间链环的结构（链接 σ_3 区和 σ_4 区的橙黄色“螺纹”状结构），橙黄色圆点表示链环的无序区域，此无序区很可能延伸到酶的活性中心（Mg^{2+}）附近。有一点是清楚的，即链环的有序部分位于 RNA 出口通道内，阻止 RNA 链的过长延伸。

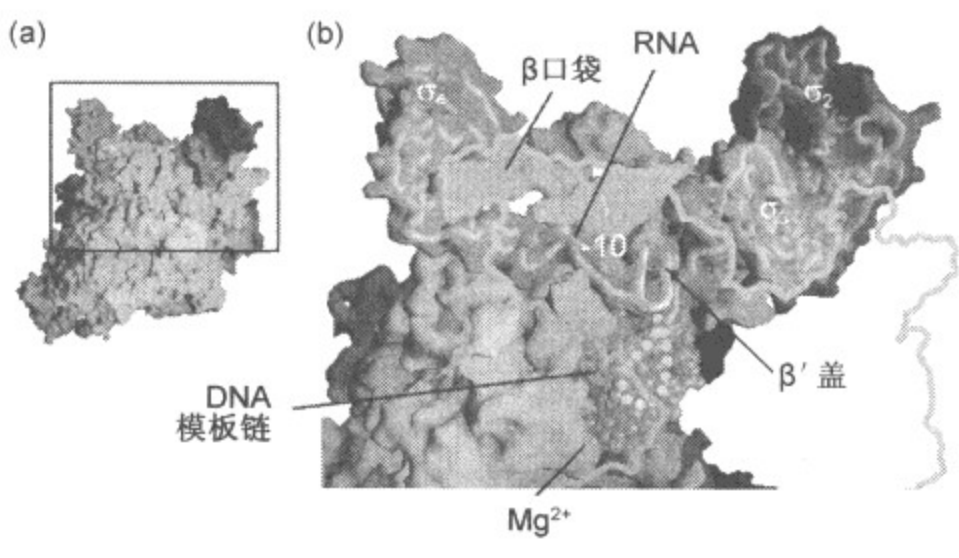


图 6.39　σ_3-σ_4 链环参与流产性转录的模型。（a）全酶结构示意图。各颜色的表示同图 6.37，只是 σ 亚基用橙色和褐色表示。（b）图（a）中加框区域的放大图。为显示被覆盖的结构，αI 及 β 亚基的部分区域被移除，分别用灰色和蓝色轮廓线表示。Σ 结构域的表面是透明的，以显示其内部的 α 碳原子骨架，σ_3、σ_4 结构域以及 σ_3-σ_4 链环用橙色表示，紫红色球体代表活性位点的 Mg^{2+}，紧邻与 DNA 模板链（绿色）互补配对的 RNA 转录物（红色）末端。RNA-DNA 杂合体用空间填充模型表示，橙色球体表示 σ_3-σ_4 链接区的无序部分靠近活性位点的可能途径。用一连串的红色球体表示 RNA 链的 5′端。如果 RNA 与链接区大小相近，则它们对排除通道的空间竞争作用是显著的。（*Source*：Murakami et al.，*Science* 296：p. 1284. © 2002 by the AAAS.）

这些发现使我们提出如下假设：位于 RNA 逐出通道内的 σ_3-σ_4 链环与初生 RNA 间存在着对此空间占据的竞争作用。通常情况下，链环能够赢得竞争，而初生 RNA 则以流产性产物形式释放出来。这种情况可能会重复发生，因为流产性转录产物在摩尔数量级上多于完整转录物（在 *E. coli* 约为 11 倍之多）。当 RNA 分子成功延伸 12nt 以上时，RNA 链将链环置换出来，使链环或与核心酶解离，或只能松散的结合在核心酶上。

为验证这一假设，Darst 及同事用 *E. coli* 全酶进行失控转录实验（第 5 章），σ 亚基分别为野生型或缺失第 503 位以后氨基酸残基的突变型（$\sigma^{70}\Delta C$），该突变 σ 缺失了 σ_3-σ_4 间的链环，根据假设，不会有过多流产性转录物，失控转录实验证实了这一推断。

最后，为什么聚合酶会如此低效地起始转录，在实现真正转录起始前，不得不反复地起始、再起始转录？其中解释之一是 σ 需要与活性位点十分靠近，以帮助起始核苷酸的结合。链环无序部分向活性位点的紧密靠近需要链环其他部分必须位于 RNA 出口通道内，而流产性转录恰是链环在该位置处的副产品。

小结　*T. aquaticus* RNA 核心聚合酶全酶的晶体结构显示，σ 亚基与 β 和 β′ 亚基间存在着广泛的接触面，推测 σ1.1 区使酶的主通道张开，以便接纳双链 DNA 模板形成闭合启动子复合体。当主通道紧紧环抱解链 DNA 形成开放启动子复合体时，$\sigma_{1.1}$ 区被逐出主通道。晶体结构还显示，σ_3-σ_4 链环位于 RNA 的出口通道内，这也许可以解释 RNA 合成起始是一个耗时过程的原因，并

产生许多流产性转录物。当转录物生长时，会为占据出口通道而与 σ_3-σ_4 链环发生强烈的竞争作用，在达到有效长度前发生中断。

全酶-DNA 复合体的结构 为得到同质（homogeneous）的全酶-DNA 复合体，Darst 及同事将 *T. aquaticus* 全酶与图 6.40 所示的叉状-接点（fork-junction）DNA 结合，该 DNA 大部分为双链结构，包含－35 框，但－10 框位于在非模板链上有一个单链突出区，起始于－11 位。该结构模拟了开放启动子复合体中启动子的结构特征，使该复合体形成类似 RP_0 的结构（RF）。

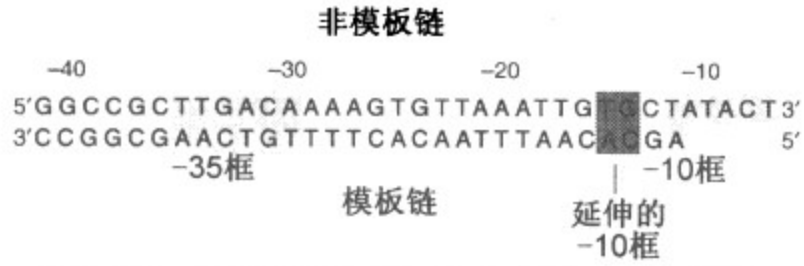


图 6.40 用于形成 RF 复合体的 DNA 结构。－35框和－10 框区为黄色阴影区，延伸－10 框区为红色阴影区。－11～－7 位的碱基为单链区，这种结构恰似开放启动子复合体中的 DNA 结构。

图 6.41 显示了全酶-启动子复合体的全貌。首先值得注意的是，从 σ 所在的位置观察，DNA 延伸穿越聚合酶的顶部。事实上，所有 DNA-蛋白质专一性相互作用都涉及 σ 亚基而不是核心酶。那么，强调 σ 亚基在转录起始中的作用也就不足为奇了。

进一步仔细观察（图 6.42）可以看出，该结构确证了以前通过生物化学和遗传学实验推测的几个特征。首先，如本章前面部分所讲到的，$\sigma_{2.4}$ 区可能参与对启动子－10 框的识别，特别是，*E. coli* σ^{70} 的 Gln_{437} 和 Thr_{440} 突变能抑制启动子区－12 位的突变，暗示这两个氨基酸与－12 位碱基存在着相互作用（图 6.21）。*E. coli* σ^{70} 的 Gln_{437} 和 Thr_{440} 对应于 *T. aquaticus* σ^A 的 Gin_{260} 和 Asn_{263}，所以我们预测这两个氨基酸靠近启动子－12 位碱基。图 6.42 证实了部分预测，Gin_{260}（绿色）确实靠近于－12 位碱基足以使两者发生接触。Asn_{263}（也用绿色标识）却因与－12 位碱基相距较远而不能发生接触，但是微小的移动就能使两者足够靠近，这种移动在体内很容易发生。

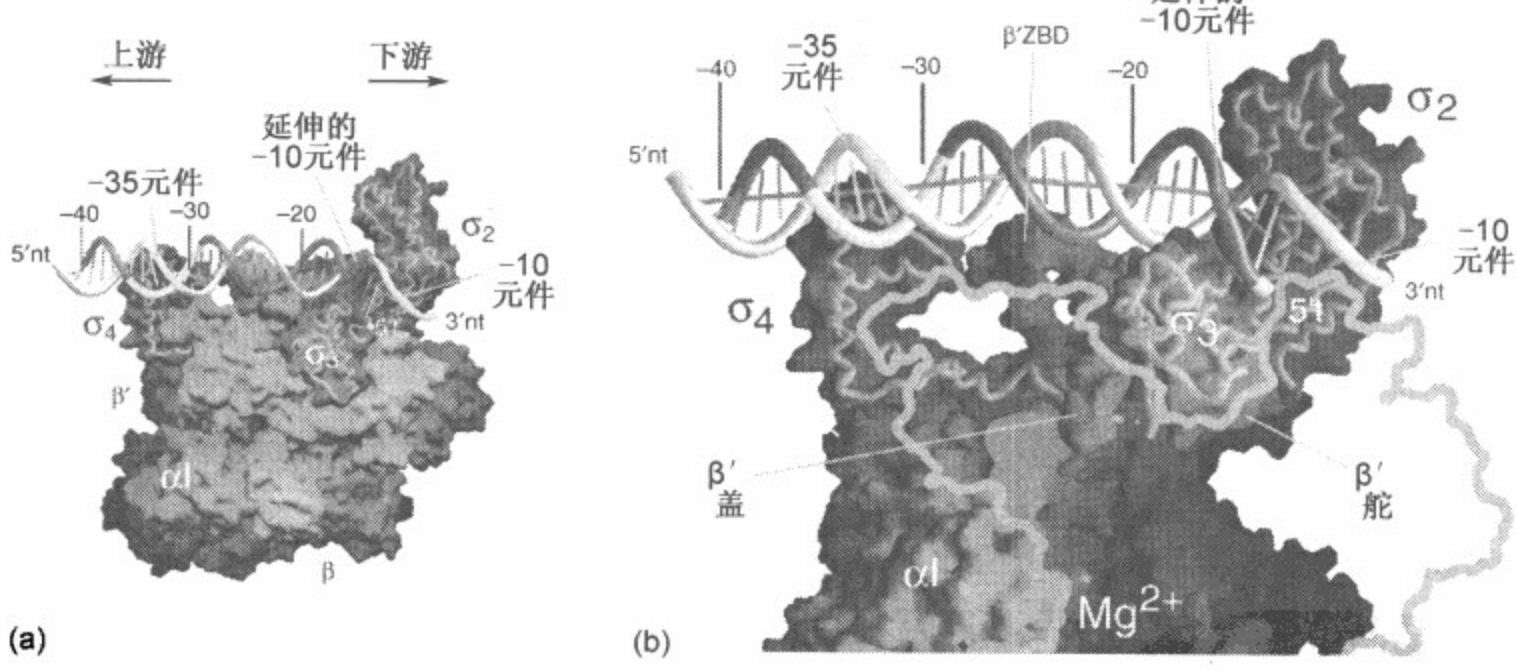


图 6.41 RF 复合体的结构。（a）复合体的整体结构。各亚基的颜色按以前常用的，扭曲梯形代表 DNA，σ 亚基的表面是局部透明的，以显示其内的 α 碳原子骨架的轮廓。（b）DNA 与蛋白质接触面的放大图示。β 亚基被移除的区域用蓝色轮廓表示。（*Source*：Murakami et al.，*Science* 296：p. 1287. © 2002 by the AAAS.）

E. coli σ^{70} 有 3 个高度保守的芳香族氨基酸残基［对应于 *T. aquaticus* σ^A 的 Phe_{248}（F_{248}）、Tyr_{253}（Y_{253}）和 Trp_{256}（W_{256}）］参与了启动子区解链。在开放启动子复合体中，这些氨基酸可能在－10 框区与非模板链结合。在 RF 复合体中，这些氨基酸（图 6.42 中黄绿色标识的氨基酸）确实与非模板单链发生相互作用。事实上，Trp_{256} 所在的位置恰好利于与碱基对 12 发生堆积，碱基对 12 是紧邻－10 框解链区的碱基对。这样，Trp_{256} 可代替－11 位的碱基对并有助于碱基对的熔解。

已知在 $\sigma_{2.2}$ 和 $\sigma_{2.3}$ 区有两个固定的碱性氨基酸残基（Arg_{237}［R_{237}］和 Lys_{241}［K_{241}］）参与结合 DNA。图 6.42 解释原因是这两个氨基酸残基（蓝色）的位置恰好有利于通过静电作用结合酸性 DNA 骨架，这种相互作用可能不具有序列特异性。

在本章的前面内容中我们了解到，$\sigma_{4.2}$ 区的特异

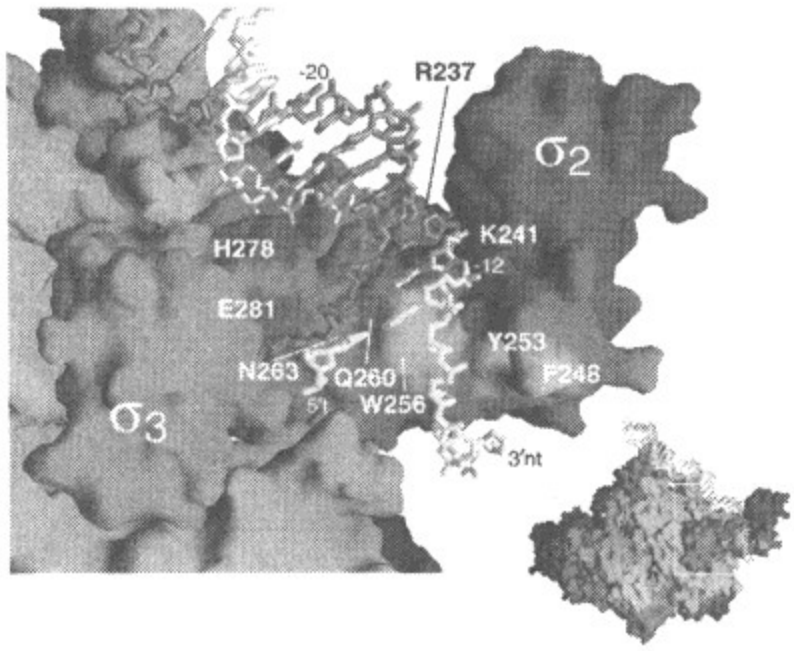


图 6.42　全酶与下游 DNA 的接触。棕色和橙色表示 σ_2 与 σ_3，通过遗传学研究，鉴定出参与下游启动子结合的氨基酸残基用其他颜色标识：识别－10 框延伸区的氨基酸残基用红色表示；识别－10 框的氨基酸残基用绿色表示；参与－10 框解链并与非模板链结合的氨基酸残基用黄绿色表示。－10 框 DNA 为黄色，延伸的－10 框 DNA 序为红色。非模板链的 3′端标记为 3′nt。在与 DNA 结合过程中起重要作用的特异氨基酸侧链也已标注。右下角小图中方框显示放大结构在 RF 复合体中的位置。（*Source*：Murakami et al.，*Science* 296：p. 1288. © 2002 by the AAAS.）

氨基酸残基有助于对启动子－35 框的结合，但令人奇怪的是，RF 结构并不能证实这些发现。特别是，－35 框与 $\sigma_{4.2}$ 相距约 6Å，而且 DNA 没有为必要的相互作用而弯曲，仍呈直线状态。因为有关－35 框与 $\sigma_{4.2}$ 相互作用的证据如此充分，所以 Darst 及同事需要解释为什么他们获得的晶体结构并没有表现出这些相互作用的特征？研究者认为，RF 结构的－35 框 DNA 因晶体堆积力而被推离了正常位置，这提示我们晶体堆积力使分子或复合体在晶体中的形态不一定与其在体内的形状相同。

基于 RF 结构和其他方面的证据，Darst 及同事提出了图 6.43a 所示的模型，解释从闭合启动子复合体向开放启动子复合体的转换，研究者也提出了 α-CTD 结合 UP 元件的假定位置。在 RP_C 中未解链的 DNA 约从－20 位置处直到下游末端的一段区域呈直线状态；但正像体内必须弯曲一样，DNA 一定会发生弯曲，从而像 RP_O 复合体那样，产生周期性 DNase 敏感性。

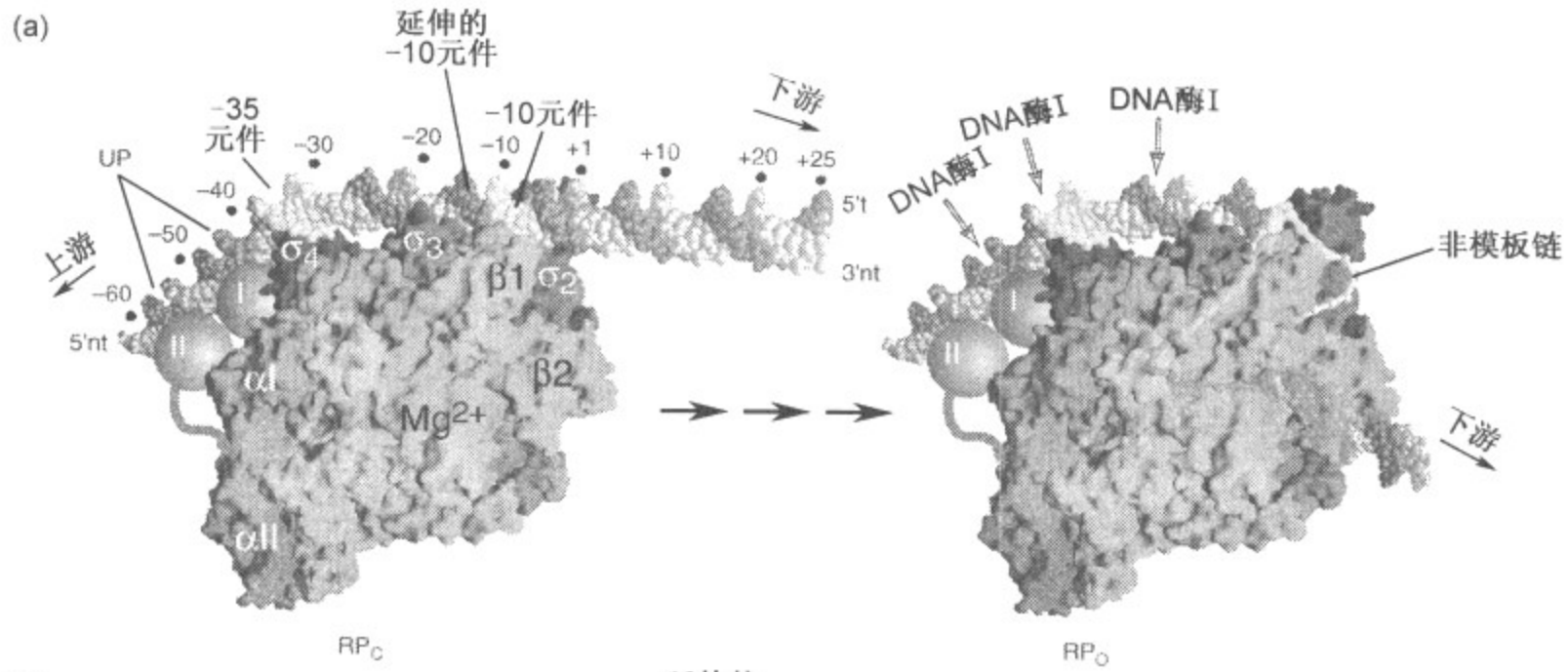
此外，就 RP_O 复合体而言，－10 框 DNA 已解链，可观察到部分非模板链。图 6.43b 是 RP_O 复合体的部分放大图，主要侧重于核心启动子-聚合酶相互作用、转录泡和下游 DNA。值得注意的是，突出于 β′亚基表面、指向 β 亚基的 β′ 舵（rudder）已从此图中移除，这个蛋白质片段跨越 β 与 β′亚基之间的缝隙，以及分开的两条单链间的缝隙，有效阻止 DNA 的复性。

Darst 及同事还有其他人员的研究结果表明，在活性中心只有一个 Mg^{2+}，但是我们认为所有 DNA 聚合酶和 RNA 聚合酶所采用的机制都需要 2 个 Mg^{2+}。为此，Dmitry Vassylyev 及同事测定了分辨率为 2.6Å 的 *T. thermophilus* 聚合酶晶体结构，非对称晶体包含有 2 个聚合酶，其中一个聚合酶含有 1 个 Mg^{2+}，另一个聚合酶有 2 个 Mg^{2+}，后者可能是 RNA 聚合酶参与 RNA 合成的形式。这两个 Mg^{2+} 为相同的 3 个天冬氨酸侧链所拥有，该侧链已结合单个 Mg^{2+}，在这互作网络中，需要邻近水分子的参与。

小结　模拟开放启动子复合体特征的 *T. thermophilus* 全酶-DNA 复合体的晶体结构揭示了以下情况。第一，DNA 主要结合在 σ 亚基上，这是启动子与 DNA 之间最为重要的相互作用；第二，预测的 $\sigma_{2.4}$ 区氨基酸与启动子－10 框的相互作用确实存在；第三，推测有 3 个高度保守的芳香族氨基酸参与了启动子解链过程，事实上，这 3 个氨基酸确实位于有助利于执行这一功能的位置上；第四，预测 σ 亚基的 2 个固定碱性氨基酸参与结合 DNA，事实上，这 2 个氨基酸确实位于有助于执行这一功能的位置上。高分辨率的晶体结构显示，聚合酶含有 2 个 Mg^{2+}，与预测的催化机制相一致。

延伸的拓扑学　当核心酶沿着 DNA 模板移动时，能否维持转录起始所产生的局部解链区？常识告诉我们，核心酶应该维持局部解链区的存在，有助于 RNA 聚合酶阅读模板链上的碱基，从而将正确的核苷酸加入到正在延伸的转录物上。实验结果证明情况确实如此。Jean-Marie Saucier 和 James Wang 将核苷酸加入到开放启动子复合体中，使聚合酶在延伸 RNA 链时能沿 DNA 向下移动。结果发现，在延伸过程中能够保持相同的熔解程度，而且，聚合酶-DNA 复合体的晶体结构清楚地显示，DNA 的两条链从全酶中两条分开的通道穿过，由此推测在延伸过程中核心酶会维持这一状态。

本章介绍的转录模型其静态特征会引起某种误导。如果将转录看做是一个动态过程，我们就可观察到 DNA 双螺旋会在移动的“转录泡”前面解旋，而又在“转录泡”的后面重新缔合。从理论上讲，RNA 聚合酶能以两种方式完成此过程，如图 6.44



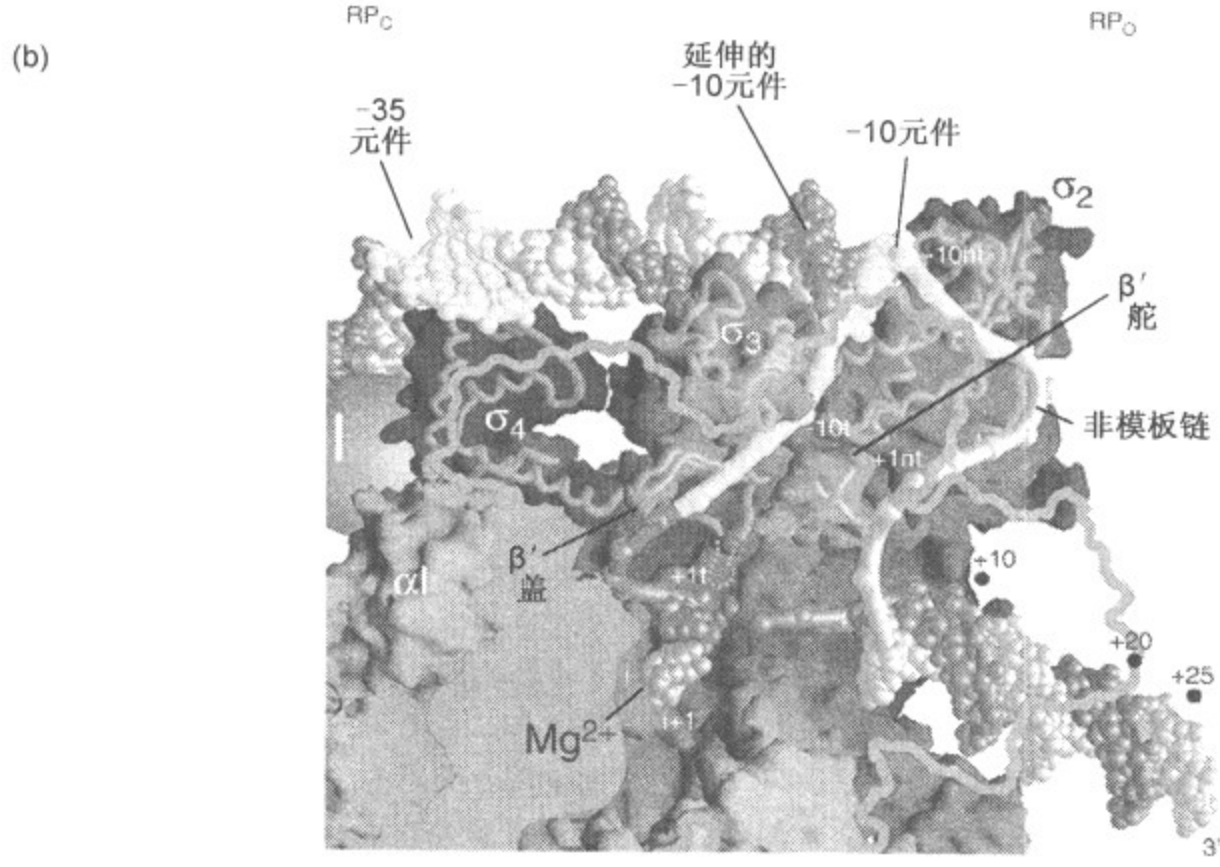


图 6.43　DNA 从−60 位至+25 位间的序列与全酶的结合模型。(a) 从闭合启动子复合体到开放启动子复合体的转换。注意由于上游 DNA 弯曲引起的 DNaseI 超敏位点，还应注意 RP_O 中 DNA 在下游解链。(b) 在开放启动子复合体的转录泡内，DNA-蛋白质相互作用的放大示意图。β 亚基的大部分没有出现在示意图中，以便观察 RP_O 中 DNA 解链情况，用青色线条勾勒出了 β 亚基缺失部分的轮廓。活性位点以 Mg^{2+} 表示，RNA 转录物的第一和第二个残基的原子分别用 i 和 i+1 表示。红色原子表示模板链的+1 位核苷酸，用+1t 标注，用+1nt 代表非模板链的+1 位核苷酸，转录泡内的模板链和非模板链用两条管线代表。β′ 亚基的舵位于解链的模板链与非模板链之间。(*Source*: Murakami et al., *Sclience* 296: p. 1288. © 2002 by the AAAS.)

所示。方式之一是在转录过程中，聚合酶和正在生长的 RNA 链围绕 DNA 模板旋转，而双螺旋 DNA 处于正常的扭曲状态（图 6.44a）。这种方式虽然能使 DNA 不发生任何扭曲，但需要大量的能量，以便使聚合酶能够不断地旋转，然后核心酶脱离转录本，使转录物不得不缠绕在 DNA 模板上，再由未知的酶解离。

另一种可能是聚合酶直线移动，前面的模板 DNA 朝一个方向旋转使双链解开，后面的模板 DNA 以相反的方向旋转重新形成双螺旋（图 6.44b）。但是，这种旋转会在 DNA 分子中引入张力。为形象地说明这个问题，想象一下电话线的解旋或者如果有电话线你不妨试一试。你可以感受到（或想象到）因电话线不断解旋而遇到的阻力，而且，如果试图将电话线缠绕得比其自然状态更紧时，也会感到阻力。确实，在解链区一端 DNA 的重旋会产生反向的补偿性扭曲而使另一端解旋，但是位于两端间的聚合酶使这种补偿不能通过解链区传递，

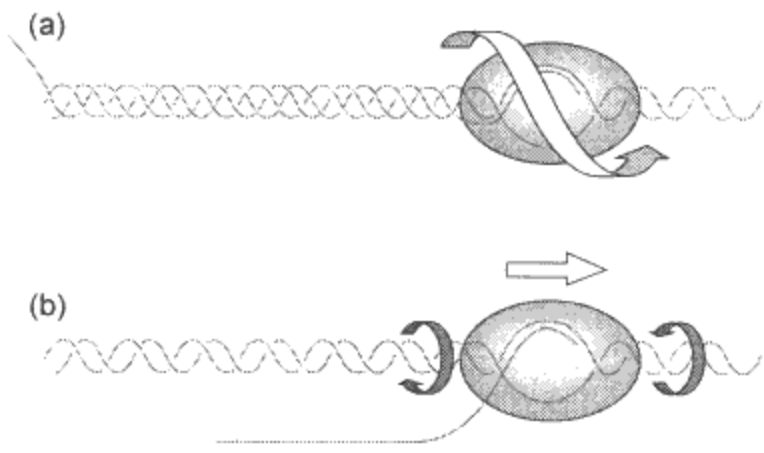


图 6.44　有关双链 DNA 转录拓扑学的两种假说。（a）RNA 聚合酶（紫色）绕双螺旋 DNA 链旋转向前移动，如黄色箭头所示，避免使 DNA 分子产生张力，但 RNA 产物（红色）也必须绕 DNA 模板旋转。（b）如黄色箭头所指，RNA 聚合酶直线移动，避免 RNA 产物（红色）绕 DNA 模板旋转，但这要求随着聚合酶的迁移，前方的 DNA 双螺旋解旋，后方的单链 DNA 重新缔合成双螺旋结构，如绿色箭头所指。由此引入的张力必须由拓扑异构酶释放。

而且环状染色体 DNA 的长跨度也隔离了解链区的两端使其不能互相补偿。

所以，如果第二种延伸机制正确有效，我们就必须解释因 DNA 解旋而产生的张力应如何释放。在第 20 章讨论 DNA 复制将了解到，拓扑异构酶（topoisomerase）能够引起 DNA 链的瞬间断裂，从而释放这种张力。我们还会看到双螺旋 DNA 扭曲产生的张力会引起螺旋的缠绕，如同扭曲的橡皮筋，这个过程称为超螺旋化（supercoiling），超螺旋化的 DNA 称为超螺旋（supercoil 或 superhelix）。因聚合酶前行而产生的解旋会引起解旋区前方的 DNA 发生补偿性过旋（compensating overwinding）（补偿性过旋是指使缠绕的电话线解旋变得困难的旋转）。习惯上，由过旋产生的超螺旋称为正超螺旋（positive supercoil）。因此，在前行聚合酶前方会产生正超螺旋；相反，在前行聚合酶后方会产生负超螺旋（negative supercoil）。该转录模型的直接证据来自拓扑异构酶突变体，该突变体不能释放超螺旋。如果突变体不能释放正超螺旋，那么正超螺旋就会在正被转录的 DNA 分子中积累。此外，不能释放超螺旋的拓扑异构酶突变体在转录过程会积聚负超螺旋。

小结　当 RNA 聚合酶沿 DNA 模板移动时，转录延伸涉及核苷酸的多聚化。伴随着聚合酶的移动，它会使 DNA 模板维持一段短的解链区。前行聚合酶前方的 DNA 解旋，而 RNA 聚合酶后面的 DNA 又重新聚合，由此在 DNA 分子上引入的张力通过拓扑异构酶加以释放。

暂停与校正　延伸过程并非始终如一，在 RNA 链延伸过程中，聚合酶时常发生暂时停顿，有时还会倒退。在 21℃ 和 1mmol/L NTP 的体外条件下，细菌系统的暂停时间非常短暂，通常只有 1～6s，但多次重复性暂停会大大延缓转录的整体速率。暂停具有重要的生理学意义，至少有两方面的原因：首先，使较慢的翻译过程能够与转录过程保持步调一致，如果翻译失败，那么暂停对于衰减作用（第 7 章）、流产性转录等十分重要；其次，暂停是转录终止的第一步，相关内容将在本章后面介绍。

有时，聚合酶会发生倒退，使生长的转录物 3′ 端伸出酶的活性位点，这不仅仅是一个放大的暂停过程。首先，该过程持续时间超过 20s，直至变为不可逆转的停止。其二，该过程只在特定条件下发生，即核苷酸浓度很低或聚合酶将错误的核苷酸加入到生长的 RNA 链中。就后一种情况而言，倒退是校正过程的一部分，这需要辅助蛋白 GreA 和 GreB 激活聚合酶固有的 RNase 活性，切除生长 RNA 链的末端，去除错配的核苷酸，重新恢复转录。GreA 切割后产生长度为 2～3nt 的短 RNA 末端片段，能够阻止但不能逆转转录停止；GreB 切割后产生长度达 18nt 的 RNA 末端片段，能够逆转转录停止。我们将在第 11 章详尽讨论真核生物的类似校正机制。

6.5　转录的终止

当 RNA 聚合酶到达基因末端的**终止子**（terminator）时，就从 DNA 模板上脱离，释放出 RNA 链。*E. coli* 细胞内有两类数目大致相同的终止子。第一类称为**内源性终止子**（intrinsic terminator），无需其他蛋白质的帮助自身就能与 RNA 聚合酶发生作用；第二类终止子依赖于辅助因子 rho（ρ），称为 rho 依赖型终止子。下面来了解一下这两种系统的转录终止机制，首先从较为简单的内源性终止子开始。

rho 非依赖型终止

rho 非依赖型（rho-independent）终止或内源性终止依赖于终止子序列中的两个元件：反向重复序列、紧随其后的一段在基因非模板链上富含 T 碱基的序列。在本节后面，我们将学习一个依赖于 RNA 转录物中“发夹”结构的终止模型，该发夹结构由 RNA 转录物中反向重复序列形成。在学习该模型之前，我们首先了解反向重复序列是如何使转录物形成发夹结构的。

反向重复序列与发夹结构　设想有以下重复序列：

5′-TACGAAGTTCGTA-3′

3′-ATGCTTCAAGCAT-5′

该序列为中心对称，其中心用圆点标出。如果在纸平面内将序列旋转 180°，并总是按 5′→ 3′方向从左向右阅读，则阅读出的序列没有任何改变。其转录物序列为：

UACGAA GUUCGUA。

它围绕其中心（下画线 G）自身互补，即自身互补碱基能够配对形成下列发夹结构：

U • A
A • U
C • G
G • C
A • U
A U
G

由于受 RNA 转角的物理限制，位于发夹顶端的 A 和 U 不能形成碱基对。

内源性终止子的结构　*E. coli trp* 操纵子（第 7 章）含有一段称为衰减子（attenuator）的 DNA 序列，可引起转录提前终止。*trp* 衰减子包含两个元件（一段反向重复序列和位于非模板链上富含 T 序列），可能是内源性终止子的关键部分。因此，Peggy Farnham 和 Terry Platt 将衰减作用作为正常终止的实验模型。

trp 衰减子的反向重复序列并不完美，但其长度仍可能达到 8bp，其中 7 个为强 G-C 碱基对，G 和 C 通过 3 个氢键而形成碱基对。反向重复序列形成的发夹结构如右：

A • U
G • C
C • G
C • G
C • G
G • C
C • G ＼
C • G ／A
U　　U
A　A

值得注意的是，U-U 和 A-A 不能形成碱基对，故在发夹结构底部形成一个小环。此外，茎右侧的一个 A 被环出（loop out），使茎长 8bp 而不是 7bp，但发夹结构仍能形成且相对稳定。

Farnham 和 Platt 提出了如下推理：当衰减子富含 T 区被转录后，DNA 模板链的 A 与 RNA 产物的 U 形成 8 个 A-U 碱基对，rU-dA 碱基对间结合非常弱，其熔解温度要比 rU-rA 或 dT-rA 碱基对的熔解温度低 20℃。由此，当 RNA 聚合酶在终止子处发生暂停时，结合较弱的 rU-dA 碱基对使 RNA 与 DNA 模板解离，导致转录终止。

有什么证据支持该模型呢？如果发夹和成串的 rU-dA 碱基对在 *trp* 衰减子中确实很重要，那么我们就可以推测，改变碱基序列、破坏任意一个碱基对都会损害衰减作用。Farnham 和 Platt 设计了如图 6.45 所示的实验方案，以此在体外分析衰减作用。研究者首先在体外转录含 *trp* 衰减子序列的 *Hpa* II 限制酶酶切片段，如果存在衰减作用，转录就会在衰减子处发生终止，产生短（140nt）的转录物；如果转录未在衰减子处终止，RNA 聚合酶继续转录直至片段的末端，则合成长度为 260nt 的失控转录物。

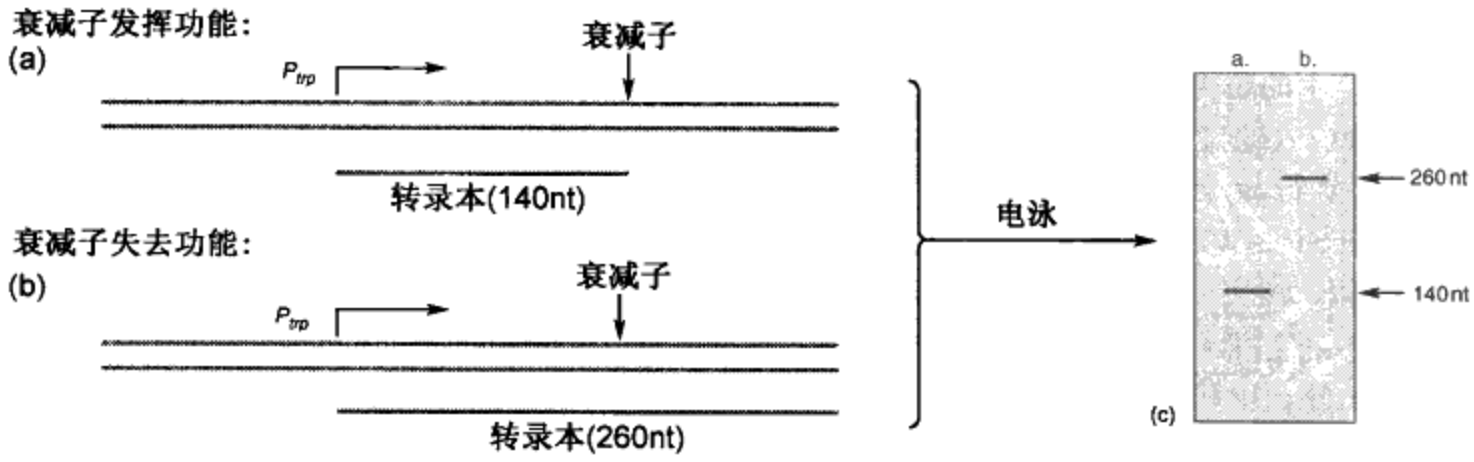


图 6.45　衰减作用分析。（a）当含有 *trp* 启动子和衰减子的 DNA 片段在衰减子发挥作用的条件下，RNA 的转录就会在衰减子处停止，产生长度为 140nt 的转录物（红色）。（b）在衰减子不发挥作用的条件下，同一 DNA 模板在相同条件下转录后，形成长度为 260nt 的失控转录物（绿色）。（c）两种不同反应的转录物很容易通过凝胶电泳加以区分，以此判断不同条件下衰减子是否会发挥功能。

通过凝胶电泳很容易将这两种转录产物分开。

当研究者把终止子非模板链上由 8 个 T 组成的串联序列更改为 TTTTGCAA 时，得到突变体 *trp* α1419，其衰减作用减弱。这与较弱的 rU-dA 碱基对是终止关键的假设相一致，因为突变体中有一半 rU-dA 碱基对被结合更强的碱基对所取代，引起衰减作用的弱化。

此外，通过体外反应，使碘-CTP 核苷酸（I-CTP）取代正常的 CTP，该突变效应可被抑制。最可能的解释是，一方面，G 与碘-C 的配对比 G 与 C 更强，使富含 GC 碱基对的发夹结构因 I-CMP 的存在而更加稳定，该效应抵消了发夹后面区域中损失的弱碱基对；另一方面，GMP 的类似物次黄鸟嘌呤核苷酸（inosin monophosphate，IMP）会减弱发夹结构内的碱基配对能力，因为 I-C 碱基对是通过 2 个氢键结合在一起的，弱于通过 3 个氢键结合在一起的 G-C 碱基对。可以确信的是，在转录反应中，ITP 取代 GTP 后可减弱衰减子的终止作用。因此，所有这些效应都与如下假设相一致：转录物的发夹结构和一连串 U 都对转录终止十分重要，然而，研究者未能分辨这些 RNA 元件在转录暂停和终止中的作用。

小结 以 *trp* 衰减子为终止子模型，Farnham 和 Platt 发现内源性终止子有两个重要特征：① 使转录物末端形成发夹结构的反向重复序列；② 非模板链上的一串 U，通过形成一串较弱的 rU-dA 碱基对而将 DNA 模板与 RNA 产物结合在一起。

终止模型 关于发夹结构和串联 rU-dA 碱基对在转录终止中的作用机制有几个假说。有两条重要的线索可以帮助我们缩小假设的范围。第一，发夹结构使人为停止（不在 rU-dA 碱基对处）的延伸复合体不稳定；第二，缺失一半重复序列的终止子即使未能形成发夹结构，仍能使转录在串联的 rU-dA 碱基对处停止。由此推导出如下一般性假设：rU-dA 碱基对引起 RNA 聚合酶的停顿，为发夹结构的形成提供了可能，而发夹结构的形成又使得 DNA 模板与 RNA 产物间原本就较弱的 rU-dA 碱基配对变得更加不稳定。稳定性降低的结果是导致 RNA 与模板分离，转录终止。

1999 年，W. S. Yarnell 和 Jeffrey Roberts 对这一假设提出了改进，如图 6.46 所示。该模型提出，当 RNA 聚合酶在终止子处发生停顿时，RNA 退出聚合酶的活性位点。可能的原因可能是由于新形成的发夹结构的帮助而将 RNA 牵引出聚合酶，也可能是由于 RNA 聚合酶继续向下游移动但并不延伸 RNA 产物，从而将 RAN 甩在其后。为了证实他们

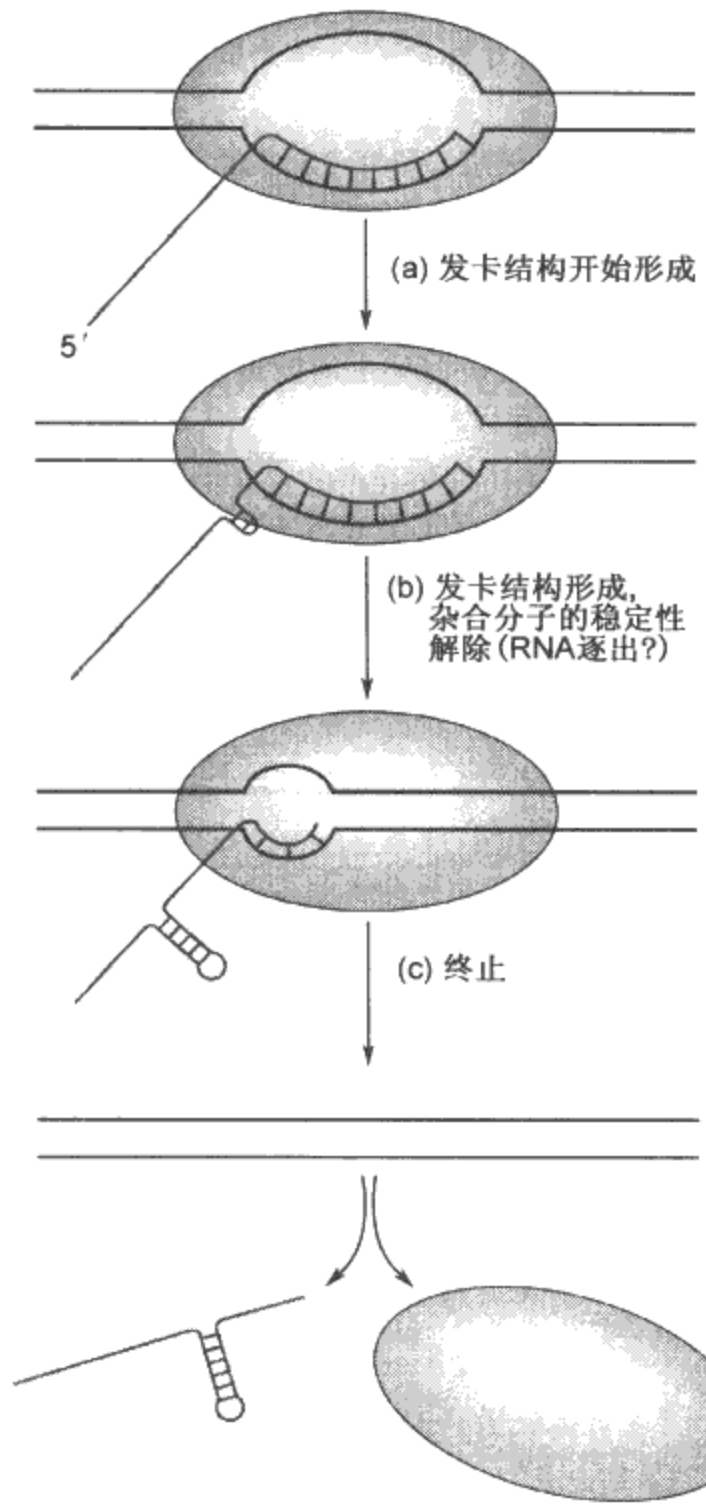


图 6.46 内源性终止子的终止模型。（a）RNA 聚合酶在连续排列的弱 rU-dA 碱基对处停顿，RNA 链在 rU-dA 碱基对的上游开始形成发夹结构。（b）当发夹结构形成后，进一步降低了 RNA-DNA 杂交分子的稳定性。不稳定性表现为以下几种形式：发夹结构的形成使转录产物退出 RNA 聚合酶，引发转录泡的解体；或相反，发夹结构的形成使转录泡解体，使 RNA 从杂合分子中释放出来。（c）RNA 产物与聚合酶完全从 DNA 模板上解离下来，转录终止。

的假设，Yarnell 和 Roberts 在实验中使用的 DNA 模板含有两个突变终止子（ΔtR2 和 Δt82），位于强启动子的下游。这些终止子在非模板链上有富含 T 的区域，但只有一半的反向重复序列，因而也就不能形成发夹结构。为补偿发夹结构的功能，研究者向反应体系中添加了能与留存的半反向重复序列互

补的寡核苷酸链，研究者的理由是，寡核苷酸能与转录本进行碱基互补配对，从而恢复发夹结构的功能。

为检验这一设想，研究者将模板吸附在磁珠上，从而可以方便地从复合物中除去模板。在适宜寡核苷酸存在或不存在条件下，用 *E. coli* RNA 聚合酶体外合成标记 RNA，最后，通过磁性吸附除去模板，获得沉淀，对沉淀和上清液进行电泳，放射自显影检测 RNA 片段。

图 6.47 显示了实验结果。泳道 1～6 无寡核苷酸，因此只有少量不完整的 RNA 产物被释放到上清液中（泳道 1、3、5 有弱 ΔtR2 和 Δt82 标记条带）。然而，在两个终止子处确实发生了停顿，特别是短时间的停顿（泳道 2、4、6 中出现较强条带），这明显表明停顿无需发夹的参与，尽管发夹对于转录本的有效释放是必需的。在泳道 7～9 中，Yarnell 和 Roberts 用寡核苷酸（t19）互补 ΔtR2 终止子留存的下游半反向重复序列，显然，如放射自显影结果所示，该寡核苷酸促进了在突变终止子处发生的终止，黑色条带表示释放到上清液中的标记 RNA，该标记 RNA 的大小与在野生型终止子处释放出的 RNA 的大小完全相同。另外，利用与△t82 终止子中留存的下游半反向重复序列互补的寡核苷酸（t18）也得到了类似的结果。

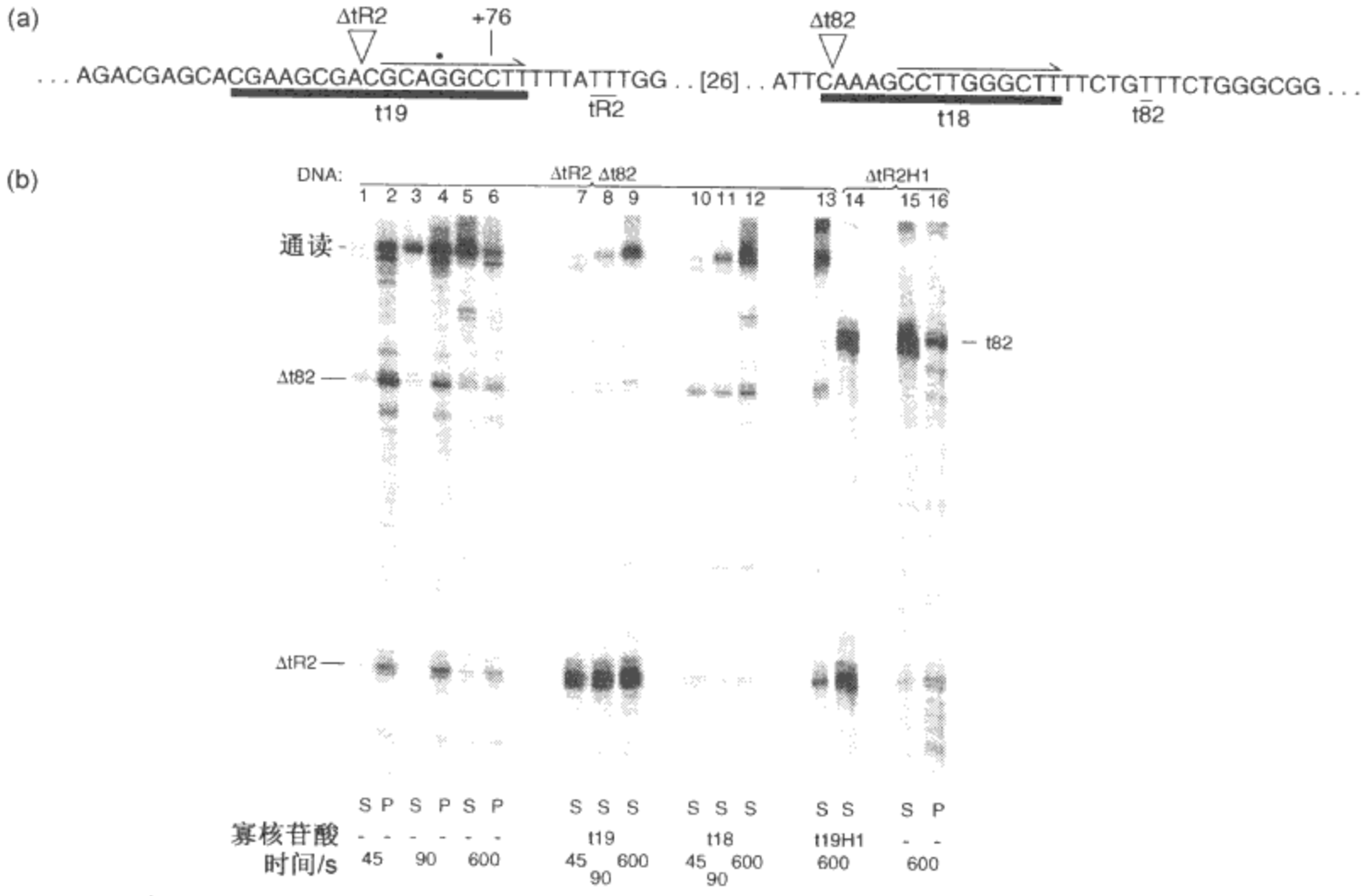


图 6.47　寡核苷酸与突变终止子的互补使转录物从延伸复合体中释放出来。(a) 实验中所用模板示意图。强启动子下游有两个突变终止子 ΔtR2 和 Δt82。两个终止子的正常终止位点用细下划线标出，粗线所示区域与寡核苷酸 t19 和 t18 互补，右向箭头表示突变终止子中保留的半反向重复序列，圆点表示 t19H1 寡核苷酸突变的碱基，作为 DNA 模板的补偿性突变。吸附到磁珠上 DNA 模板通过离心很容易从溶液中去除。(b) 实验结果。Yarnell 和 Roberts 以图 (a) 所示模板合成标记 RNA，泳道 1～6、15 和 16 无寡核苷酸；泳道 7～9 有 t19 寡核苷酸；泳道 10～12 有 t18 寡核苷酸；泳道 13～14 有 t19H1 寡核苷酸。按图底部所示时间进行转录，然后离心除去模板和及所吸附的 RNA，对沉淀 (P) 及上清 (S)（底部标示）中的标记 RNA 进行电泳和放射自显影，失控转录物及在终止子 ΔtR2 和 Δt82 处终止的转录物的位置如图左侧所示。(*Source*: (a-b) Yarnell, W. S. and Roberts, J. W. Mechanism of intrinsic transcription termination and antitermination. *Science* 284 (23 April 1999) 611-12. © AAAS.)

为进一步验证寡核苷酸与半反向重复序列之间碱基配对的重要性，研究者将寡核苷酸 t19 中的一个碱基突变而得到寡核苷酸 t19H1。泳道 13 显示，寡核苷酸的这一改变引起了 ΔtR2 终止作用的显著降低。然后对 ΔtR2 进行补偿突变，再次检验 t19H1 的作用。泳道 14 显示补偿突变恢复了 ΔtR2 的强终止作用，由于该模板含有野生型 Δt82 终止子，所以，也有大量终止在此发生。泳道 15 和 16 是无寡核苷酸 t19H1 的负对照。正如预期的那样，在终止子 ΔtR2 处很少发生终止。

综上所述，这些研究结果表明，发夹结构本身并不是终止所必需的，转录产物与下游半重复序列之间的碱基配对是发生终止所必需的，通过这种碱基配对来破坏 RNA-DNA 杂合分子的稳定性。而且，如果人为地使转录过程缓慢进行，那么，富含 T 的区域也不是必需的。Yarnell 和 Roberts 使聚合酶停留在一个既不位于反向重复序列也不位于富含 T 区域内的位点处，通过洗除核苷酸而确保聚合酶停留在此位点。然后，研究者向反应体系中加入能与人为停留位点上游序列杂交的寡核苷酸，在这些条件下，观察初生 RNA 的释放。

NusA 蛋白也能促进终止，该蛋白质能够促进发夹结构在终止子处形成。2001 年，Ivan Gusarow 和 Evgeny Nudler 指出了这个模型的实质：转录本在形成发夹结构序列的上游具有部分核心聚合酶的结合位点，即上游结合位点（upstream binding site，UBS），当 RNA 聚合酶与 UBS 结合后，会减缓发夹结构的形成，不利于转录终止。而 NusA 蛋白能够松弛 RNA 聚合酶与 UBS 间的结合，因此，NusA 蛋白促进发夹结构的形成，因而也就具有促进转录终止的功能。在第 8 章中，我们将详细讨论 NusA 蛋白及其发挥功能的模式，了解上述模型的证据。

小结 细菌终止子的特性使其具有两方面的作用：① 某种方式的碱基配对会降低 RNA-DNA 杂合分子的稳定性；②并由此引起转录停顿。正常的内源性终止子通过在转录物内形成发夹而满足第一个条件，通过在发夹下游形成一连串的 U 而满足第二个条件。

rho 依赖型终止

Jeffrey Roberts 发现蛋白质 rho 能显著降低 RNA 聚合酶体外转录某些噬菌体 DNA 的能力，抑制的结果导致转录终止。rho 无论何时引发转录终止，RNA 聚合酶都必须重新起始转录，由于转录起始是一个耗时过程，因此发生的净转录（net transcription）很少。为确定 rho 蛋白是一个真正的终止因子，Roberts 开展了以下实验。

rho 影响链的延伸但不影响起始 如同 Travers 和 Burgess 用［γ-^{32}P］ATP 和［^{14}C］ATP 分别检测 RNA 合成的起始和总 RNA 的合成一样，Roberts 用［γ-^{32}P］GTP 和［^{3}H］UTP 进行相同目的的实验。在逐渐增加 rho 浓度的情况下，Roberts 用这两种核苷酸进行体外转录反应，图 6.48 是实验结果。可以看到 rho 对转录起始几乎没有影响，如果有影响，则转录的起始速率就会上升，但 rho 却引起总 RNA 合成的显著下降。这与 rho 因子终止转录、重新启动耗时的起始过程的观点一致。这一假说推测 rho 会引起较短转录物的合成。

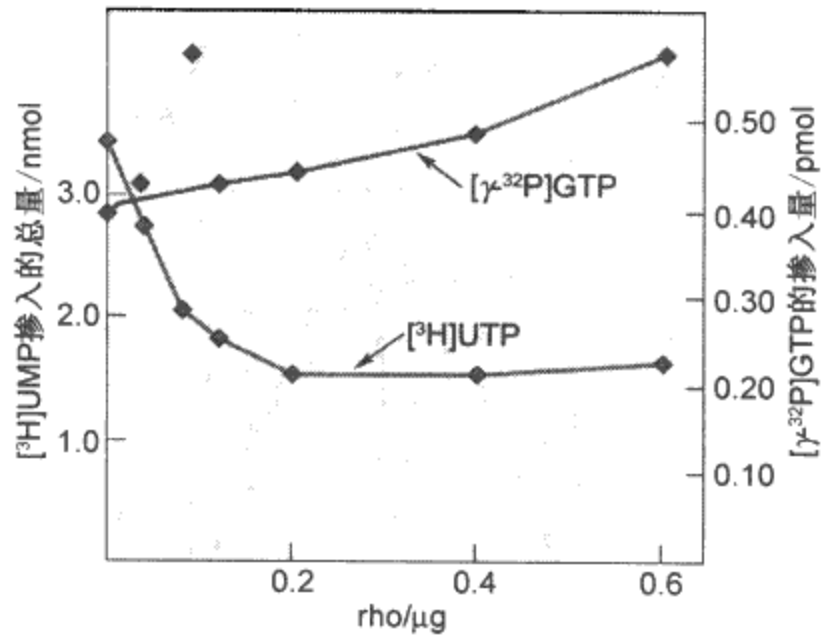


图 6.48 rho 能降低 RNA 合成的净合成率。在不断增加 rho 因子浓度条件下，Roberts 用 *E. coli* RNA 聚合酶转录 λ 噬菌体 DNA，用［γ-^{32}P］GTP 检测转录起始（红色），用［^{3}H］UTP 检测转录延伸（绿色）。Rho 因子降低了转录延伸的速率，但不影响转录起始。（*Source*：Adapted from Roberts, J. W. Termination factor for RNA synthesis, *Nature* 224：1168-74，1969.）

rho 引起短转录物的合成 用凝胶电泳或 1969 年 Roberts 在实验中使用的超速离心法可相对容易地测定 RNA 转录物的大小，但只根据短转录物的存在还不足以得出 rho 引起转录终止的结论，因为 rho 也许只是一种 RNase，可将长的转录物切割成短片段。

为排除 rho 只有核酸酶活性的可能性，Roberts 首先在无 rho 情况下用^{3}H 标记 λ RNA，然后将相对较大的 RNA 片段加入到含 rho 的新反应体系中，^{14}C-UTP 作为标记 RNA 前体。最后用超速离心法测量^{14}C-和 ^{3}H-标记的 λ RNA 的大小。图 6.49 显示了研究结果。连续曲线显示^{3}H- RNA 没有区别，即使在第二个反应中与 rho 因子共孵育，其大小也没有变化，说明 rho 没有核酸酶活性。但是，在 rho 存在情况下合成的^{14}C-RNA（图 6.49b 中的红色曲

线）明显短于无 rho 时合成的^{14}C-RNA。所以，rho 能够引起较短 RNA 的合成，再次证明了 rho 在转录终止中的作用。没有 rho 因子，转录物会延伸到非正常的长度。

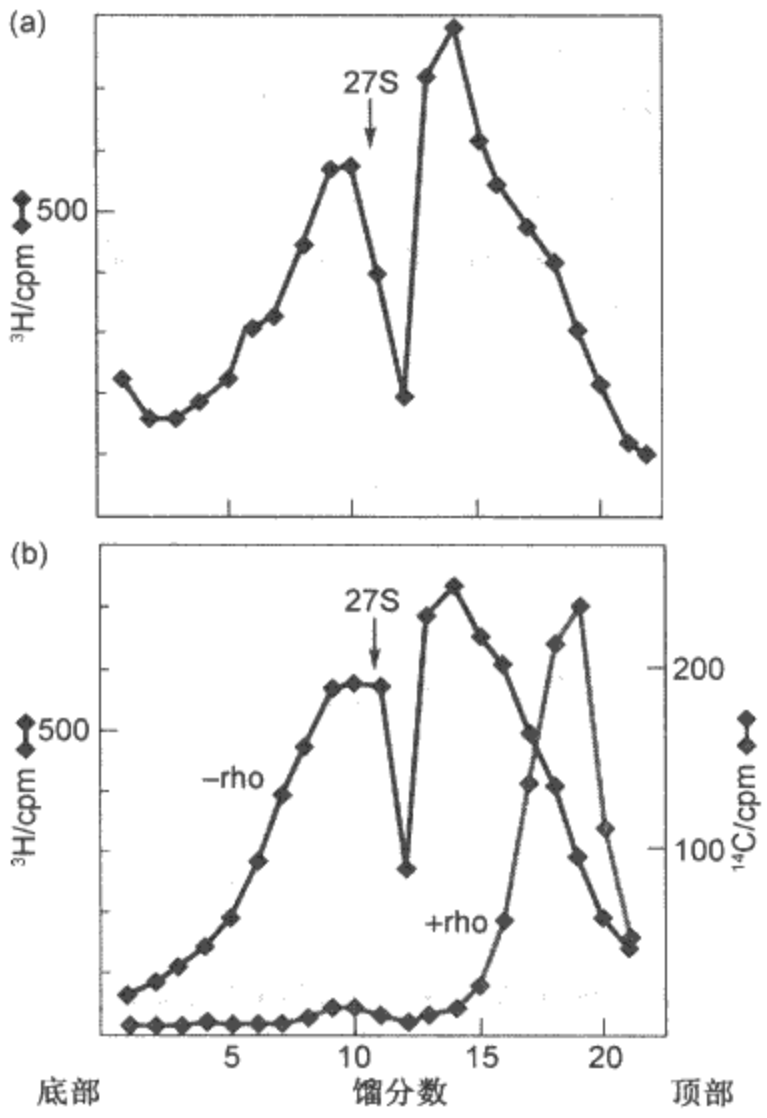


图 6.49　rho 因子能够降低 RNA 产物的大小。（a）在 rho 因子不存在的条件下，Roberts 用 *E. coli* RNA 聚合酶转录 λ DNA，反应体系中添加［^{3}H］UTP，以标记 RNA 产物。反应结束后，进行超速离心，使产物依据其大小而分离。Roberts 收集离心管底部的片段，在图左侧，长度最大 RNA 产物其放射性强度较低。（b）在 rho 因子存在的条件下，Roberts 用 *E. coli* RNA 聚合酶转录 λ DNA，反应体系中添加［^{14}C］UTP，以标记 RNA 产物。同时添加（a）实验中获得的［^{3}H］UTP 标记的 RNA 产物。反应结束后，进行超速离心，使产物依据其大小而分离。在 rho 因子存在的条件下，^{14}C 标记的 RNA 产物（红色）其条带靠近离心管的顶部（右侧），其长度相对较小。在 rho 因子不存在条件下，获得的^{3}H 标记的 RNA 产物（蓝色）其长度相对较大，且没有发生任何变化。结果证明：rho 对已合成的 RNA 的长度不会产生影响，但它能使正在合成的转录物的长度缩短。（*Source*：Adapted from Roberts, J. W. Termination factor for RNA synthesis, *Nature* 224: 1168-74, 1969.）

rho 从 DNA 模板上释放转产物　最后，Roberts 用超速离心法分析了 rho 有或无条件下 RNA 产物的沉降性质。没有 rho 时，RNA 产物与 DNA 模板共沉淀（图 6.50a），表明该转录物未从与 DNA 结合的复合体中释放。相反，有 rho 时，转录物（图 6.50b）的沉降速率很低，不与 DNA 模板共沉淀。所以，rho 似乎能够将 RNA 从 DNA 模板上释放出来。事实上，rho（希腊字母 ρ）代表“释放”。

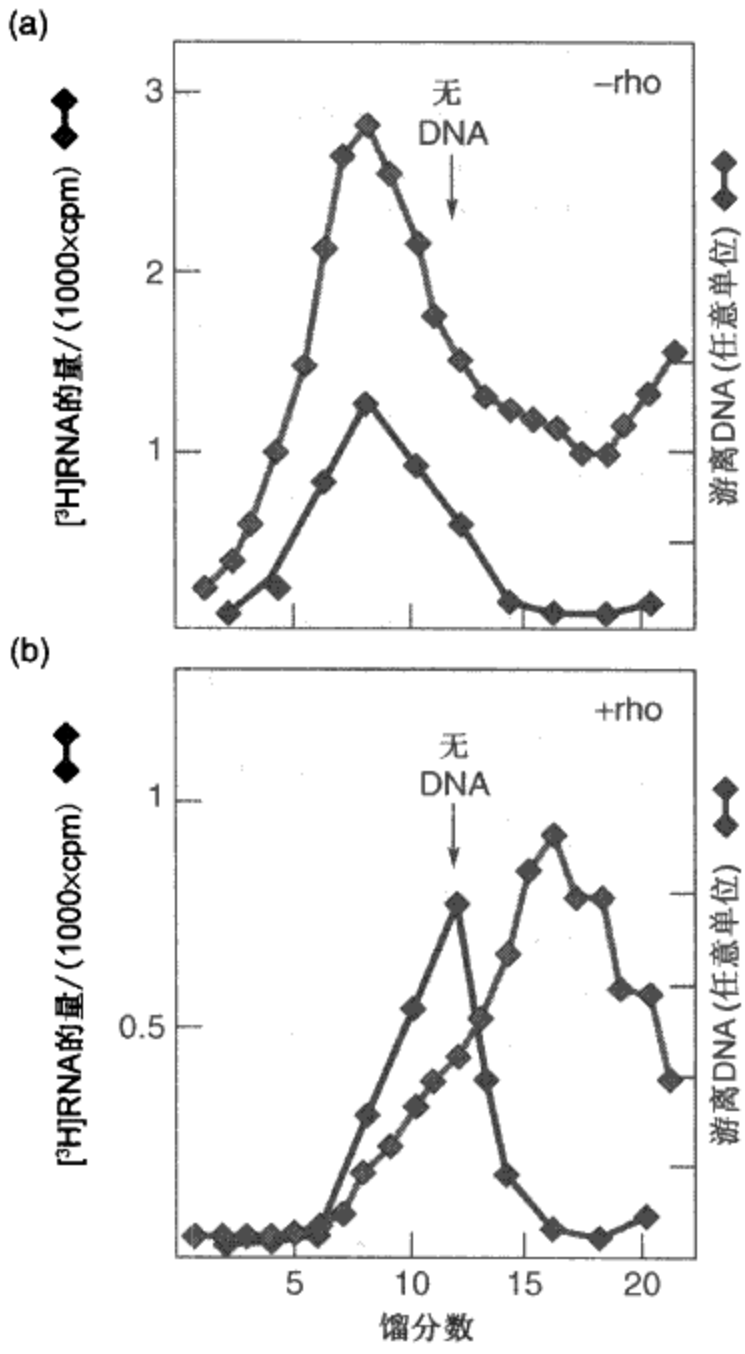


图 6.50　rho 因子从 DNA 模板上释放 RNA 产物。在图 6.49 所示的相同实验条件下，Roberts 在 rho 因子存在（a）或不存在（b）两种情况下，于体外转录 λDNA，然后通过超速离心方法分析^{3}H 标记的转录产物（红色）是否与 DNA 模板（蓝色）结合在一起。（a）在 rho 不存在时合成的转录物会与 DNA 模板共沉淀，该复合体比游离的 DNA 大；（b）在 rho 存在时合成的转录物不与 DNA 模板共沉淀，其所在的位置对应于相对较小的分子。有 rho 因子的转录能够从 DNA 模板上释放出转录物。（*Source*：Adapted from Roberts, J. W. Termination factor for RNA synthesis, *Nature* 224: 1168-74, 1969.）

rho 因子的作用机制　rho 因子如何执行其功能？我们已经知道 σ 因子通过与核心酶紧密结合而而在 RNA 聚合酶中发挥作用。因而可以设想，rho 可能以同样的方式发挥作用。但 rho 对 RNA 聚合酶没有亲和性。σ 是聚合酶全酶的关键部分，而 rho 因子不是。

取而代之的是（图 6.51），rho 在转录物终止位点上游的 rho 装载位点（rho loading site）处与转录物结合，该位点长 60～100nt，富含胞嘧啶残基，相比较而言无二级结构。相同的 6 个亚基组成 rho 六聚体（hexamer），每个亚基均具有 ATPase 活性，rho 与 RNA 结合后激活了其 ATPase 活性，进而提供能量，推动 rho 六聚体沿着 5′→3′方向移动，追赶 RNA 聚合酶。直到 RNA 聚合酶在终止子处延宕，此时转录物已形成发夹结构，RNA 聚合酶在终止子处的延宕使 rho 追赶上来，释放出转录物。在 1987 年，Terry Platt 及同事研究表明，rho 具有 RNA-DNA 解旋酶活性。所以，当 rho 与停留在终止子处的 RNA 聚合酶相遇后，就可在转录泡内将 RNA-DNA 解链，从而释放出转录物，实现转录终止。

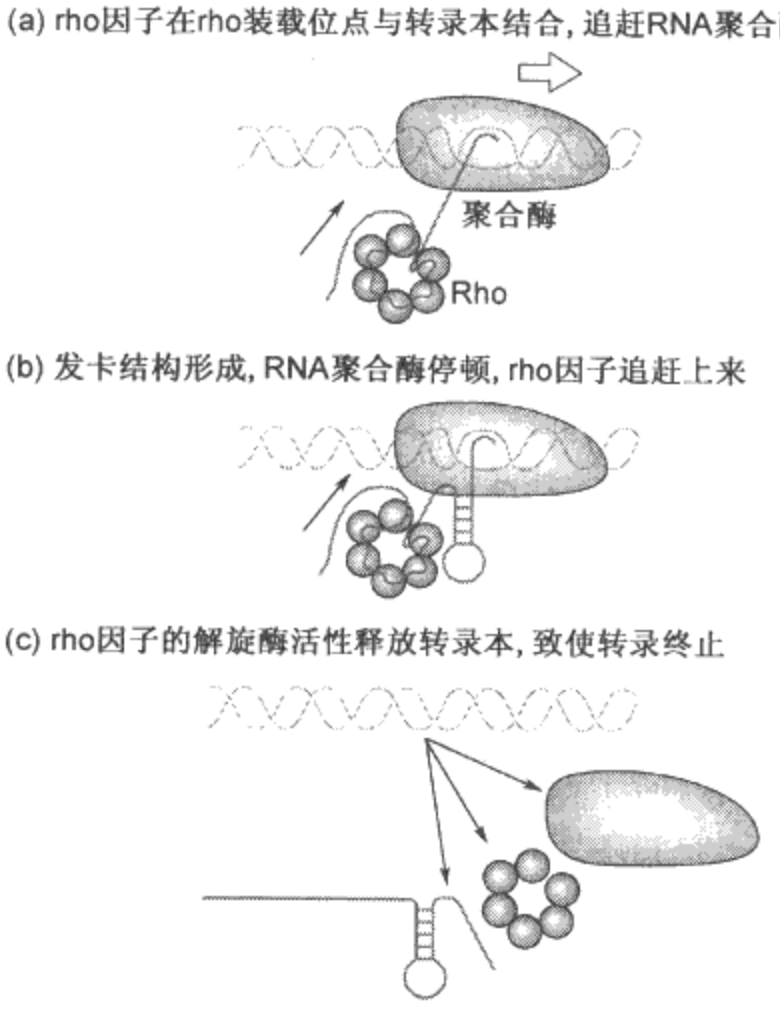


图 6.51　rho 依赖型终止模型。（a）当 RNA 聚合酶（红色）还在进行 RNA 的合成时，rho 因子（绿色）在 rho 装载位点处与转录本结合，沿转录本生长方向追赶聚合酶。（b）当转录本形成发夹结构时，引发 RNA 聚合酶的暂停，为 rho 因子追赶上聚合酶提供了机会。（c）rho 因子的解旋酶活性使 RNA-DNA 双链杂合体结构解离，从而释放出转录本。

终止。

rho 六聚体如同一个锁紧的垫圈，呈开环状，末端具有分支，可允许生长的 RNA 链通过开口进入六聚体。rho 的 6 个亚基均具有 RNA 结合位点，使 RNA 酶像拉链一样从一个亚基滑向另一个亚基。

小结　rho 依赖型终止子含有反向重复序列，能使转录物形成发夹结构，该类终止子无成串的 T。rho 结合在正在生长的转录本上，当 RNA 聚合酶在发夹处发生停顿时，rho 追赶上 RNA 聚合酶，通过使 RNA-DNA 杂交分子解旋而使转录本从 DNA-聚合酶复合体中释放出来，转录终止。

总结

RNA 聚合酶是转录过程中的催化剂。*E. coli* 的 RNA 聚合酶由核心酶和 σ 因子组成，核心酶拥有基本转录装置，σ 因子指导核心酶转录特异的基因。σ 因子通过使 RNA 聚合酶全酶与启动子紧密结合而引发转录起始，依赖 σ 因子的紧密结合需要在转录起始位点附近发生 DNA 局部解链，解链长度为 10～17bp，从而形成开放启动子复合体。由于只能指导全酶与特定启动子的结合，所以 σ 因子能够选择将要被转录的基因。转录起始过程持续到约有 9nt 或 10nt 添加到 RNA 产物上，然后，σ 因子或与核心酶解离，或松散地继续结合在核心酶上，核心酶转变为延伸特异性构象，离开启动子，执行延伸过程。σ 因子可以被不同的核心酶重复利用，控制聚合酶对利福平敏感性或抗性的是核心酶，而不是 σ 因子。

原核生物启动子含有两个区域，其中心分别位于转录起始位点上游的 −10bp 和 −35bp 处。在 *E. coli*中，这两个区域的保守序列分别为 TATAAT 和 TTGACA。一般来讲，启动子内的相应区域越接近这些保守序列，启动子起始转录的能力就越强，某些特别强的启动子在核心启动子上游具有额外的元件（UP 元件），使启动子对 RNA 聚合酶具有更强的吸引力。

σ 因子之间具有 4 个相似的区域，其中 2.4 亚区和 4.2 亚区分别参与了聚合酶对−10 框和−35 框的识别。

核心酶 β 亚基位于 RNA 聚合酶催化磷酸二酯键形成的活性位点附近。在转录起始阶段，σ 因子也位于活性位点附近。模板转换实验确定了两个与聚合酶相互作用的 DNA 位点，一个是弱结合位点，包括 DNA 的解链区和位于或靠近聚合酶 β 亚基的催化位点，在此位点发生的蛋白质-DNA 的相互作用主要依赖于静电作用，具有盐敏感性；另一个是强结合位点，涉及活性位点的下游 DNA 和聚合酶的 β

和β′亚基。α亚基具有独立折叠的N端功能域和C端功能域。C端功能域能够识别启动子的UP元件，并与之结合，可实现聚合酶与启动子间的紧密结合。

转录延伸过程实质是，当RNA核心聚合酶沿着模板DNA前行时，进行核苷酸的聚合作用。转录泡的长度为10～18bp，包含长度约9bp的RNA-DAN杂合分子。转录泡的移动需要聚合酶前方DNA的解旋、后方DNA的重新缔合，这个过程会将张力引入DNA模板链中，需通过拓扑异构酶释放。

嗜热水生菌RNA核心聚合酶的晶体结构如同一只蟹钳，催化中心位于通道内，含有1个能与3个Asp残基配位的Mg^{2+}，该通道能够引导DNA穿过聚合酶。

T. aquaticus RNA核心聚合酶全酶的晶体结构显示，σ亚基与β和β′亚基间存在着广泛的接触面，推测$\sigma 1.1$区使酶的主通道张开，以便接纳双链DNA模板形成闭合启动子复合体。当主通道紧紧环抱解链DNA形成开放启动子复合体时，$\sigma_{1.1}$区被逐出主通道。晶体结构还显示，σ_3-σ_4链环位于RNA的出口通道内，这也许可以解释RNA合成起始是一个耗时过程的原因，并产生许多流产性转录物。当转录物生长时，会为占据出口通道而与σ_3-σ_4链环发生强烈的竞争作用，在达到有效长度前发生中断。

嗜热水生菌聚合酶全酶-DNA复合体的晶体结构与开放复合体结构相似，可以归纳得出以下结论：①DNA主要结合在σ亚基上；②预测的$\sigma_{2.4}$区氨基酸与启动子－10框的相互作用确实存在；③3个高度保守的芳香族氨基酸参与启动子熔解的推测是真实的，在正确位置的这3个氨基酸确实发挥了这样的作用；④σ亚基上2个不变的碱性氨基酸参与对DNA结合的推测是真实的。高分辨率的晶体结构显示，聚合酶含有2个Mg^{2+}，与可能的催化机制相一致。

内源性终止子具有两个重要元件：① 使转录物末端形成发夹结构的反向重复序列，该发夹结构可以破坏RNA-DNA杂合分子的稳定性；② 非模板链上的一串U，通过形成一串较弱的rU-dA碱基对而将DNA模板与RNA产物结合在一起。这些元件共同作用，引起RNA聚合酶的停顿，释放出转录产物。rho依赖型终止子含有反向重复序列，能使转录物形成发夹结构，该类终止子无成串的T。rho结合在正在生长的转录本上，当RNA聚合酶在发夹处发生停顿时，rho追赶上RNA聚合酶，通过使RNA-DNA杂交分子解旋而使转录本从DNA-聚合酶复合体中释放出来，转录终止。

复习题

1. 解释下列结论：①RNA聚合酶核心酶只能微弱地转录完整的T4噬菌体DNA，而全酶却能够很好地转录该模板；②同全酶一样，核心聚合酶也能高效地转录小牛胸腺DNA。

2. Bautz及同事是如何证明*E. coli* RNA聚合酶全酶只能特异地转录T4噬菌体的极早期基因，而核心酶却不具有这样的特异性？

3. Bautz及同事是如何证明全酶以非对称方式转录T4噬菌体DNA，而核心酶能以对称方式转录T4噬菌体DNA？

4. 阐述测量蛋白质与DNA最紧密结合复合体解离速率的实验策略，举例说明松散结合与紧密结合的实验结果。这些结果和核心酶与全酶，以及含启动子DNA序列的结合有何关系？

5. 温度对聚合酶-启动子复合体的解离速率有何影响？能够暗示复合体的什么性质？

6. 图示说明闭合复合体与开放复合体的区别。

7. 图示典型的原核启动子及含UP元件的启动子，不要求写出精确的序列。

8. 描述实验并给出结果，证明*E. coli* RNA聚合酶合成流产性转录物。

9. 图表说明*E. coli*转录起始过程中的4个步骤。

10. 描述实验并给出结果，检测σ因子对转录起始速率和转录延伸速率的影响。

11. 如何证明σ因子确实不会加速转录延伸的速率？

12. 根据上述两个问题的实验结果，能够得出的最终结论是什么？

13. 描述实验并给出结果，证明σ因子可重复利用。在同一张图中，显示证明核心聚合酶决定利福平抗性的实验结果。

14. 图示说明“σ循环”，假设在延伸阶段σ因子从核心酶上脱离下来。

15. 描述荧光共振能量转移实验结果，证明σ因子在延伸阶段并没有与核心酶解离。

16. 如何修正σ循环才能解释上述问题的实验结果？引述一个纯化延伸复合体支持修正σ循环的实验。

17. 描述实验并给出结果，证明当RNA聚合酶与启动子结合后，会引起哪些碱基对的解离，解释该过程的机制。

18. 描述实验并给出结果，估计*E. coli* RNA聚合酶在转录期间所解离的碱基对数目。

19. σ因子的什么区域参与了对启动子①－10

框和②－35 框的识别？不需要特异残基的命名。说明这些结论的遗传学证据。

20. 描述能够为 σ 因子 4.2 亚区与启动子－35 框之间存在的相互作用提供生物化学证据的结合分析实验方案。

21. 引述支持 *E. coli* RNA 聚合酶 α 亚基参与识别启动子 UP 元件这一假设的实验证据。

22. 描述如何利用限制性蛋白水解酶来确定蛋白质的结构域，如 *E. coli* RNA 聚合酶 α 亚基。

23. 描述确定负责对利福平和链霉溶菌素抗性或敏感性的聚合酶亚基的实验方案。

24. 描述实验并给出结果，证明 *E. coli* RNA 聚合酶 β 亚基位于形成磷酸二酯键的活性位点附近。

25. 描述模板转换实验，证明 *E. coli* RNA 聚合酶 β 亚基参与 DNA 的弱结合，而 β′亚基参与 DNA 的强结合，并阐述实验结果。

26. 描述 RNA-DNA 交联实验，证明转录延伸复合体中存在 RNA-DNA 双链杂合体，其长度至少为 8bp。

27. 基于 X 线晶体图像，简图显示 RNA 核心聚合酶的大致结构，指出亚基、催化中心和利福平结合位点的大致位置。根据该结构，提出利福平抑制转录的机制。

28. 基于 X 线晶体图像，简图显示 RNA 聚合酶全酶的大致结构，指出 σ 因子 2、3 和 4 结构域的大致位置。

29. 全酶的晶体结构没有包括 σ 因子的 1.1 区，但其他证据表明 $\sigma_{1.1}$ 定位闭合和开放启动子复合体内，这些被建议的位置是什么？至少引述 2 个证据支持你提出的定位。

30. 基于全酶的晶体结构，提出一个能够解释在转录起始时发生流产的原因。按照该模型，σ 因子的哪一部分对流产转录负责？解释这种无效起始模式的原因。

31. 简图说明开放启动子中全酶-DNA 复合体的晶体结构，集中说明全酶与 DNA 间的相互作用。在与 DNA 结合过程中，哪个亚基发挥了最大作用？

32. 已知 σ 因子的 2.4 亚区和 4.2 亚区分别参与了对启动子－10 框和－35 框的识别，全酶-DNA 复合体的晶体结构证实了这个模型的哪一部分？解释没有证实另一部分的原因。

33. 提出当 RNA 聚合酶沿着 DNA 模板移动时维持 DNA 熔解泡的两种方式。哪一种方式获得了证据的支持？用一两句话列举证据。

34. 内源性终止子的两个重要元件是什么？如何获知它们的重要性（列举证据）。

35. 提出在内源性终止子处停顿不需要发夹结构的证据。

36. 提出证据，证明停顿位点上游 RNA 的某些碱基配对是内源性终止子所必需的。

37. rho 依赖型终止子看上去像什么？在这类终止子中，rho 的作用是什么？

38. 如何证明 rho 会降低 RNA 的净合成而不会减少链的起始？描述实验方案并阐述结果。

39. 描述实验并给出结果，证明存在时会形成短转录产物。该实验亦应证明 rho 并不是简单地行使核酸酶功能。

40. 描述实验并给出结果，证明 rho 因子能够将转录本从 DNA 模板上释放出来。

分析题

1. 图示具有 10bp 茎和 5nt 环的 RNA 发夹结构。写出形成此发夹结构的序列，展示该序列的线性结构和发夹结构。

2. 包含 σ^{70} 的 *E. coli* RNA 聚合酶全酶识别的一个 *E. coli* 启动子－10 框的非模板序列为 5′-CAT-AGT-3′。（a）在第一位置发生的 C→T 突变是上升突变还是下降突变？（b）在最后位置发生的 T→A 突变是上升突变还是下降突变？解释你的回答。

3. 你为研究 *E. coli* 基因转录终止而开展实验，对该基因 3′端测序后获得如下结果：

5′-CGAAGCGCCG**ATTGC**CGGCGCTTTTTTTT-3′

3′-GCTTCGCGGC**TAACG**GCCGCGAAAAAAAA-5′

然后你通过对此序列的改变而获得突变基因，具体的序列改变如下：

突变体 A：

CGAAACTAAG**ATTGC**AGCAGTTTTTTTTT

突变体 B：

CGAAGCGCCG**TAGGA**CGGCGCTTTTTTTT

突变体 C：

CGAAGCGCCG**ATTGC**CGGCGCTTACGGCC

你对每一个突变基因都进行了分析，以测定转录的终止，并获得了如下结果：

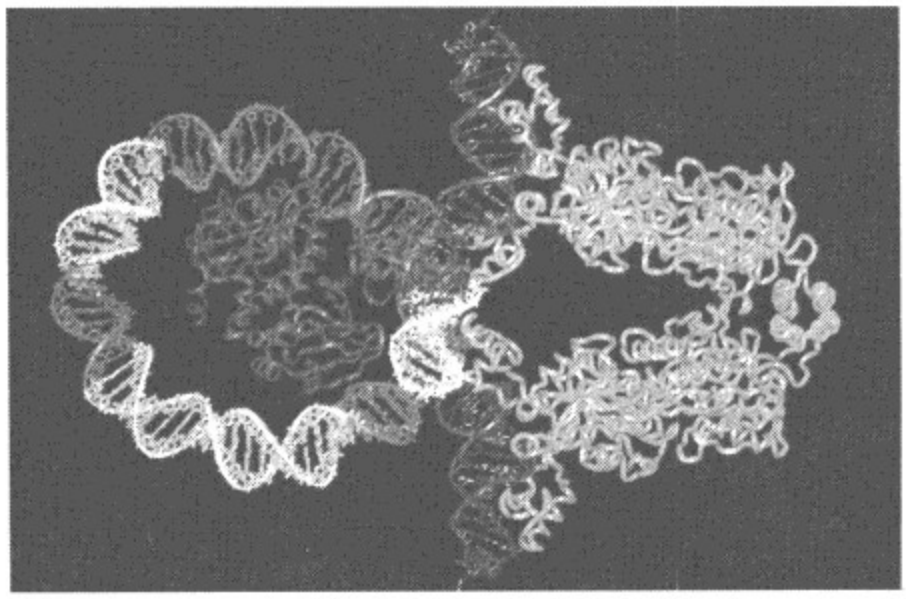
测试的突变基因	无 rho	有 rho
野生型基因	100%终止	100%终止
突变体 A	40%终止	40%终止
突变体 B	95%终止	95%终止
突变体 C	20%终止	80%终止

a. 绘图表示上述野生型基因序列转录出 RNA 分子的结构。

b. 尽可能全面地解释实验结果。

翻译　徐启江　校译　李玉花　许志茹

第7章　操纵子：细菌转录的精细调控



***lac* 阻遏物（紫色）与 CAP-cAMP 二聚体（蓝色）和 DNA 结合的计算机模拟图。**
© Mitchell Lewis/SPL/Photo Researchers, Inc.

E. coli 基因组含有 3000 多个基因，其中一些基因一直处于激活状态，因为其产物是恒定需求的。而有些基因在多数时间内是关闭的，因为很少需要其产物。例如，当最适能源葡萄糖缺乏而只有阿拉伯糖时，才需要代谢阿拉伯糖的酶，但这种情况并不多见，所以编码这些酶的基因通常是关闭的。为什么细胞不让所有基因都开启，以便在需要时能够快速提供相应的酶呢？原因在于基因表达是一个高代价的过程，需消耗大量能量产生 RNA 和蛋白质。如果 *E. coli* 细胞的所有基因一直都处于表达状态，合成 RNA 和蛋白质势必消耗细胞太多的能量，使 *E. coli* 不能与更高效的生物体竞争。因此，基因表达调控对生命活动至为关键。本章将探讨细菌调控其基因表达的策略，即将功能相关基因集合成组而便于表达调控，这样一组彼此相邻、协同调控的基因串称为操纵子（operon）。

7.1　*lac* 操纵子

发现的第一个操纵子已成为理解操纵子概念的典型范例。该操纵子含有 3 个基因，编码的酶可使 *E. coil* 细胞利用**乳糖**（lactose），因此称为 ***lac* 操纵子**（*lac* operon）。如果将 *E. coil* 细胞培养在既含**葡萄糖**（glucose）又含乳糖的培养基上（图 7.1），细胞会耗尽葡萄糖，停止生长。它们能否通过自身调整去适应新糖源？短期内显然不能，但是经过约 1h 的生长停滞期后生长就恢复了。在停滞期细胞启动了乳糖操纵子的表达，开始积累参与乳糖代谢的酶。图 7.1 所示的生长曲线称为“二次生长”（diauxic）曲线，源于拉丁文“*auxilium*”，寓意为帮助，因为两种糖源均有助于 *E. coli* 细胞生长。

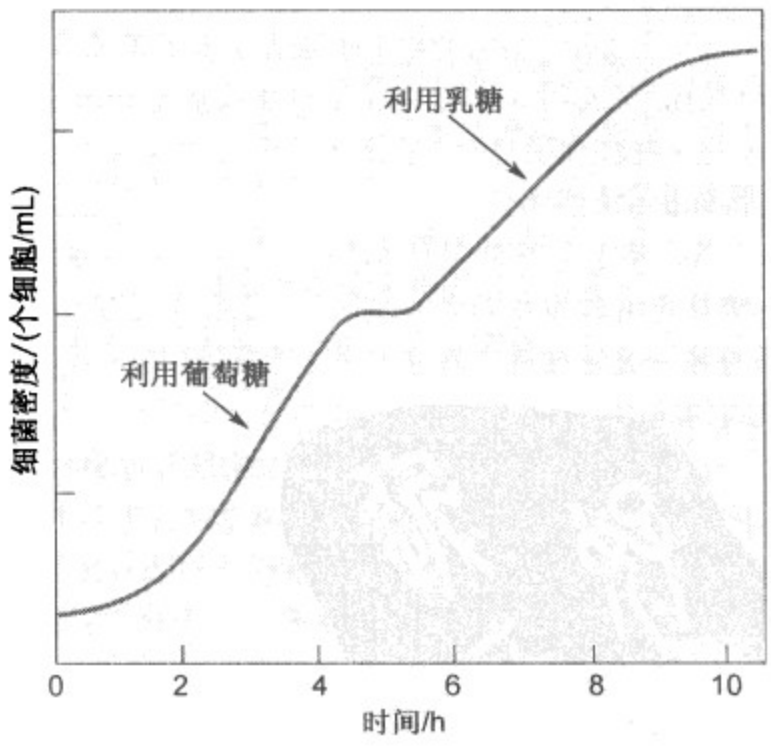


图 7.1　二次生长曲线。 *E. coli* 细胞在含有葡萄糖和乳糖的培养基中培养，用细菌密度（个细胞/mL）对时间（h）作图。细胞利用葡萄糖进行快速生长，直到耗尽葡萄糖。细胞诱导乳糖代谢酶时，生长停止；当所诱导的酶出现之后，生长恢复。

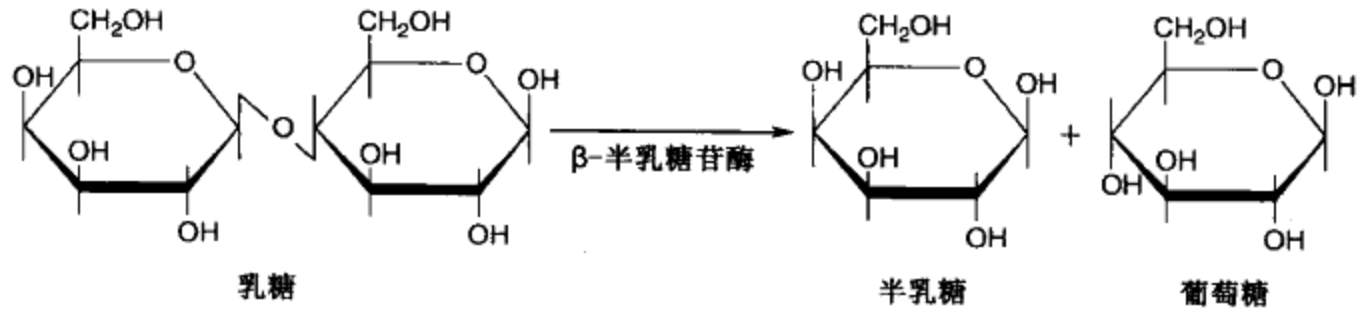


图 7.2　β-半乳糖苷酶反应。β-半乳糖苷酶，断裂半乳糖（紫色）和葡萄糖（蓝色）之间的 β-半乳糖苷键（灰色），从而使二糖乳糖分解为半乳糖和葡萄糖。

这些酶都是什么呢？首先，*E. coli* 需要一种酶将乳糖运输到细胞中，这种酶就是半乳糖苷透性酶（galactoside permease）。其次，细胞需要一种将乳糖降解成半乳糖和葡萄糖的酶。图 7.2 显示该反应过程。由于乳糖由两种单糖组成，所以称为二糖（disaccharide），半乳糖和葡萄糖这两个六碳糖通过 β-半乳糖苷键连接在一起，因此乳糖又叫做 **β-半乳糖苷**（β-galactoside），将其分解为半乳糖和葡萄糖的酶叫做 β-半乳糖苷酶（β-galactosidase）。编码半乳糖苷透性酶和 β-半乳糖苷酶的基因在乳糖操纵子中并列在一起，同时还伴有编码**半乳糖苷乙酰转移酶**（galactose transacetylase）的另一个结构基因，该酶在乳糖代谢中的功能还不清楚。

编码乳糖代谢 3 种酶的基因按以下顺序成簇排列：β-半乳糖苷酶（***lacZ***）、半乳糖苷透性酶（***lacY***）、半乳糖苷乙酰转移酶（***lacA***）。它们由同一个启动子起始转录，产生一条**多顺反子信使**（polycistronic message）mRNA 分子，因此，通过控制该启动子就可同时调控这 3 个基因。多顺反子一词源于**顺反子**（cistron），顺反子是“基因”的同义词。所以，多顺反子信使 RNA 是指该信使 RNA 含有一个以上基因的信息。在 mRNA 上每个顺反子都有自己的核糖体结合位点，所以，每个顺反子能够独立地被核糖体翻译而合成相应的蛋白质。

如本章开头所提到的，乳糖操纵子（与其他许多操纵子一样）是受细胞严格控制的，事实上，对操纵子的调控存在两种系统。首先是**负调控**（negative control），就像汽车的制动器一样，打开制动器才能开动汽车。***lac* 阻遏物**（*lac* repressor）就是负调控的“制动器”，只要缺乏乳糖，*lac* 阻遏物就使操纵子处于关闭（或阻遏）状态。这是一个经济有效的机制，否则细胞合成缺乏底物的酶无疑是一种浪费。

如果负调控像汽车制动器，那么**正调控**（positive control）就像汽车的加速器踏板。在 *lac* 操纵子中，从操纵基因上移除阻遏物（松开制动器）不足以激活乳糖操纵子，还需要正调控因子**激活子**（activator）的参与。我们将会了解到激活子通过激活乳糖操纵子的转录而对低水平葡萄糖做出响应，而高水平葡萄糖使激活子浓度保持较低水平，不能激活操纵子的转录。这种正调控系统的优点在于高浓度葡萄糖使操纵子几乎处于关闭状态。如果没有葡萄糖水平响应途径，乳糖单独存在就足以激活操纵子，但在葡萄糖还处于可利用水平时就激活 *lac* 操纵子的表达是不适宜的，因为 *E. coli* 细胞代谢葡萄糖比代谢乳糖更容易。因此，在有葡萄糖时激活乳糖操纵子对 *E. coli* 细胞是一种浪费。

小结　*E. coli* 的乳糖代谢由两种酶执行，可能还有第三种酶参与。编码这三种酶的基因成簇地排列在一起，从同一启动子处一起转录，产生一条多顺反子信使。这 3 种基因在功能上相关，所以在表达上也相关，它们一起被关闭或被开启。负调控保证了乳糖操纵子在乳糖缺乏条件下处于关闭状态；而正调控保证了乳糖操纵子在葡萄糖存在时，即使有乳糖存在也仍处于相对失活状态。

lac 操纵子的负调控

图 7.3 举例说明了 *lac* 操纵子调控的一个方面：负调控的经典模式。虽然本章后面内容和第 9 章对于这种经典调控模式的介绍过于简化，但有助于在初学时对操纵子概念的理解。负调控（negative control）意味着操纵子一直处于开启状态，除非在某种物质的干扰下使其关闭。关闭乳糖操纵子的“某种物质”是 *lac* 阻遏物。*lac* 阻遏物是调节基因 ***lacI*** 的产物（如图 7.3 最左边所示），是由 4 种相同多肽组成的四聚体，结合在启动子右侧的**操纵基因**（operator）上。一旦阻遏物结合操纵基因，操纵子就被抑制（repressed）。因为操纵基因与启动子毗邻，阻遏物结合到操纵基因上阻碍了 RNA 聚合酶结合启动子以及转录操纵子。由于基因不能转录，所以操纵子处于关闭或抑制状态。

只要没有乳糖可利用，乳糖操纵子就处于抑制状态。另一方面，当所有葡萄糖耗尽而乳糖存在时，会有一种移除阻遏物的机制，使操纵子解除阻遏，并利用新的营养物质。这种机制是如何运作的呢？阻遏物是一种**变构蛋白**（allosteric protein）（希腊语中 *allos* 的意思是“其他的”，*steros* 的意思是

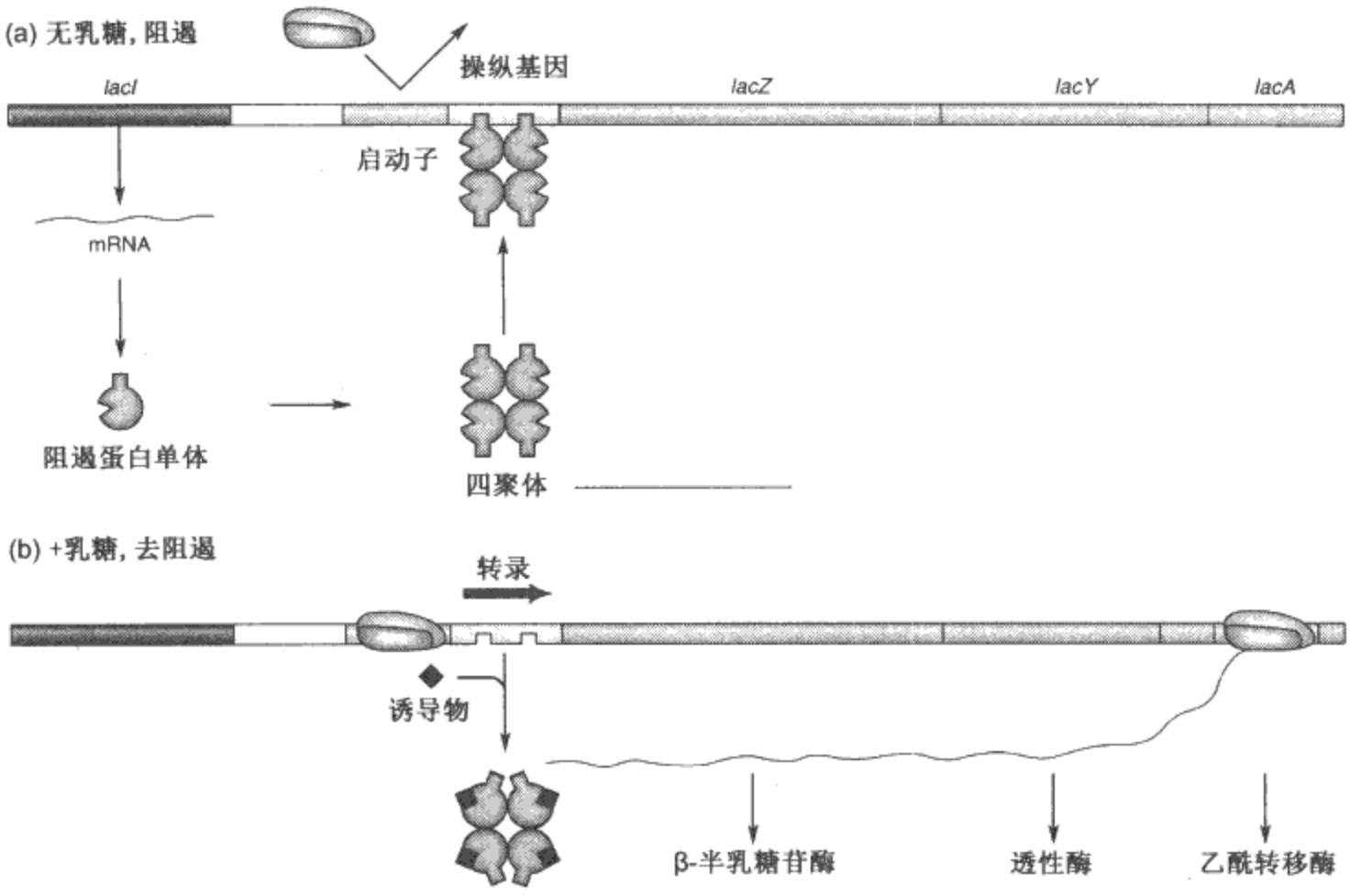


图 7.3 *lac* 操纵子负调控。（a）无乳糖时，操纵子阻遏。*lacI* 基因表达产生阻遏物（绿色），阻遏物结合操纵基因，阻止 RNA 聚合酶转录乳糖操纵子基因。（b）有乳糖时，操纵子去阻遏。诱导物（黑色）与阻遏物结合，改变阻遏物构象，使其不再与操纵基因结合。阻遏物脱离操纵基因，RNA 聚合酶起始结构基因的转录，产生多顺反子 mRNA，进而翻译产生半乳糖苷透性酶、β-半乳糖苷酶和半乳糖苷乙酰转移酶。

"形状"）。一种分子与变构蛋白的结合会改变该蛋白质另一位点的构型，从而改变它与第二种分子的相互作用。第一种分子称为乳糖操纵子**诱导物**（inducer），因为该分子结合阻遏物，引起阻遏物蛋白质构象改变，使阻遏物与操纵基因（第二个分子）解离，从而诱导了操纵子的表达（图 7.3b）。

诱导物的本质是什么呢？它实际上是另一种形式——**异乳糖**（allolactose）。β-半乳糖苷酶将乳糖降解为半乳糖和葡萄糖的同时，催化部分乳糖重排形成异乳糖。如图 7.4 所示，在异乳糖中半乳糖和葡萄糖连接方式不同于乳糖（乳糖是 β-1,4-糖苷键连接，异乳糖是 β-1,6-糖苷键连接）。

也许你会提出这样的问题，如果 *lac* 操纵子是被抑制的，没有透性酶将乳糖运到细胞中，没有 β-半乳糖苷酶催化反应的发生，乳糖是如何代谢生成异乳糖的？答案在于阻遏的遗漏和操纵子极低的本底水平表达。由于每个细胞仅有 10 个左右的四聚体阻遏物，因此不需要太多的诱导物就可以启动 *lac* 操纵子的表达。随着 *lac* 操纵子产物的增多，会引起更多诱导物的形成，就像滚雪球一样。

操纵子的发现

Francois Jacob、Jacques Monod 及同事对操纵子概念的发展是融合遗传学和生物化学分析的一个成功典范。1940 年，Monod 研究 *E. coli* 乳糖代谢的可诱导性时发现，β-半乳糖苷酶是乳糖代谢的主要特征，可被乳糖及其他半乳糖苷所诱导。Monod 与 Melvin Cobn 利用 β-半乳糖苷酶抗体检测 β-半乳糖

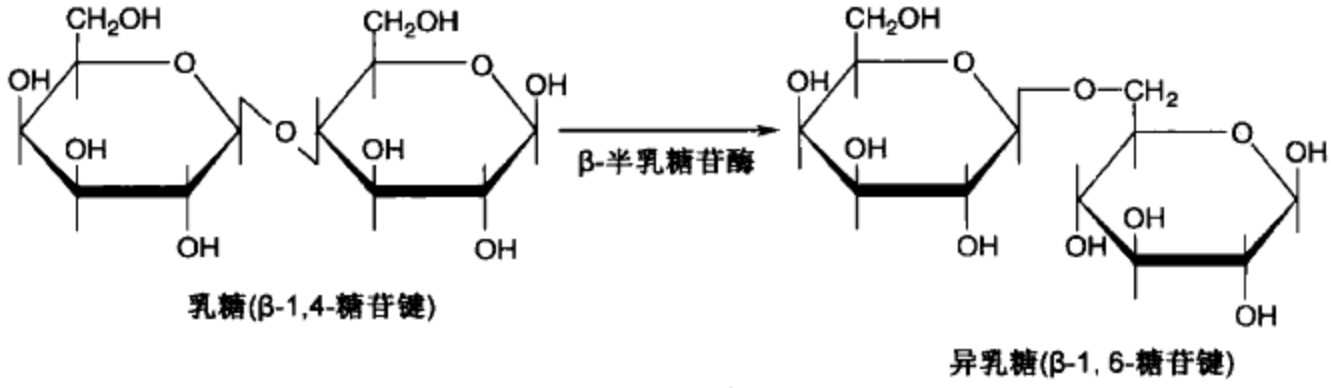


图 7.4 乳糖与异乳糖的转换。在 β-半乳糖苷酶催化下，乳糖分子重排形成诱导分子异乳糖。注意，糖苷键由 β-1,4 变为 β-1,6。

苷酶蛋白时发现，该蛋白质的量在诱导过程中是逐渐增加的。因为应答乳糖的过程中产生了较多的基因产物，显然，β-半乳糖苷酶基因自身可被诱导表达。

为剖析其复杂性，研究者发现了一些突变体[最初称为“隐蔽突变体”（cryptic mutant）]，这些突变体能合成β-半乳糖苷酶，但不能在含乳糖的培养基上生长。这些突变体缺失了什么成分？为了弄清原因，Monod 及同事向培养有野生型及突变型 *E. coli* 的培养基内添加放射性半乳糖，发现野生型和突变体细胞均不能吸收半乳糖，即使突变体处于诱导条件下也不能吸收半乳糖，而诱导的野生型细胞却能积累半乳糖苷。该实验揭示了两个事实。首先，在野生型细胞中，某种物质（半乳糖苷透性酶）与β-半乳糖苷酶协同被诱导，并负责将半乳糖苷转运到细胞内。其次，突变体编码这种蛋白质的基因似乎是存在缺陷的（Y^-）（表 7.1）。

表 7.1　隐蔽突变体（*lacY*$^-$）对半乳糖苷积累的影响

基因型	诱导物	半乳糖苷的积累
Z^+Y^+	−	−
Z^+Y^+	+	+
Z^+Y^-（隐蔽突变）	−	−
Z^+Y^-（隐蔽突变）	+	−

Monod 将这种物质命名为半乳糖苷透性酶，由于在尚未分离出的蛋白质之前就为其命名，Monod 一直承受着同事们批评。后来他评论说：“这好比两位传统的英国绅士，即使他们彼此孰知对方的姓名和声望，但在他人正式引见之前彼此却不能说话”。在努力纯化半乳糖苷透性酶的过程中，Monod 及同事又鉴定出一种蛋白质，即半乳糖苷乙酰转移酶，该酶与β-半乳糖苷酶和半乳糖苷透性酶协同被诱导。

到了 20 世纪 50 年代后期，Monod 了解到这 3 种酶的活性（因而推测可能是 3 个基因）能够同时被半乳糖苷所诱导，同时也发现一些无须诱导的**组成型突变体**（constitutive mutant），它们总能够合成 3 个基因的产物。Monod 意识到遗传学分析将会极大地推进研究的深入，因此他开始与巴斯德研究所的 Francois Jacob 合作，继续开展相关研究。

通过与 Arthur Pardee 合作，Jacob 和 Monod 获得了**部分二倍体**（merodiploid）（部分二倍体细菌），既携带野生型（可诱导）等位基因，又携带组成型等位基因。其中可诱导等位基因为显性，表明野生型细胞产生一种物质使 *lac* 基因保持关闭，直到被诱导表达。由于该物质均能关闭来自组成型亲本和可诱导型亲本的基因，因此使部分二倍体也可诱导。当然，该物质就是 *lac* 阻遏物。组成型突变体编码阻遏物的基因（*lacI*）是缺陷的，所以这些突变体为 *lacI*$^-$（图 7.5a）。

阻遏物的存在需要一些特殊的 DNA 序列，以便阻遏物与之结合。Jacob 和 Monod 将这些特异 DNA 序列称为**操纵基因**（operator）。这种相互作用的特异性表明可进行遗传突变分析，即操纵基因的突变可消除它与阻遏物的相互作用，这些突变也应该是组成型的。那么，如何区分操纵基因的组成型突变与阻遏物基因的组成型突变呢？

他们认为通过确定突变体是显性突变还是隐性突变即可区分二者。Jacob 用类比法来解释其观点，他把部分二倍体细菌的 2 个操纵基因比作 2 个无线电接收器，分别控制着房间的 2 个门。部分二倍体的 2 个阻遏物基因就像 2 个无线电发射器，各自发出相同的信号使房门保持关闭。其中一个阻遏物基因突变就像一个无线电发射器损坏一样，而另一个无线电发射器仍然发挥功能，两个房门仍保持关闭。也就是说，部分二倍体的 2 个 *lac* 操纵子仍然处于被阻遏状态，所以这种突变应该是隐性突变（图 7.5a），我们已经观察到确实如此。

另一方面，部分二倍体的一个操纵基因突变就像一个接收器损坏了一样，不能接收无线电信号，该房间门则处于打开状态。而另一个房门的接收器功能正常，所以仍然保持关闭。也就是说，部分二倍体的突变型 *lac* 操纵子始终是去阻遏的，而野生型 *lac* 操纵子保持阻遏。因此，该突变是显性突变，但只针对受突变操纵基因调控的 *lac* 操纵子（图 7.5b），这种突变称为**顺式显性**（*cis*-dominant）突变（拉丁语“*cis*”顺式意指“在这里”），即部分二倍体中只对同 DNA 上的基因表现显性效应，而对其他 DNA 的基因无显性效应。Jacob 和 Monod 确实发现了这样的顺式显性突变体，并证明了操纵基因的存在。携带组成型操纵基因（operator constitutive）的突变体称为 O^c。

如果阻遏物基因发生突变使阻遏物失去对诱导物应答的能力，这种突变体的 *lac* 操纵子调控情况又将如何？这类突变会导致 *lac* 操纵子的非诱导性，表现为顺式（*in cis*）或反式（*in trans*）显性效应（对部分二倍体中不同 DNA 分子具有调控作用，拉丁语“*trans* 反式”意指“在对面”），因为突变体阻遏物即使在诱导物或野生型阻遏物存在时，仍保持与 2 个操纵基因的结合（图 7.5c），类似于无法关闭的无线电信号。Monod 及同事发现了 2 个这类突变体，随后 Suzanne Bourgeois 又发现了更多其他突变体。将这类突变体为命名为 I^s，与组成型阻遏物突变体（I^-）相区别，I^- 使阻遏物不能与操纵基因结合。

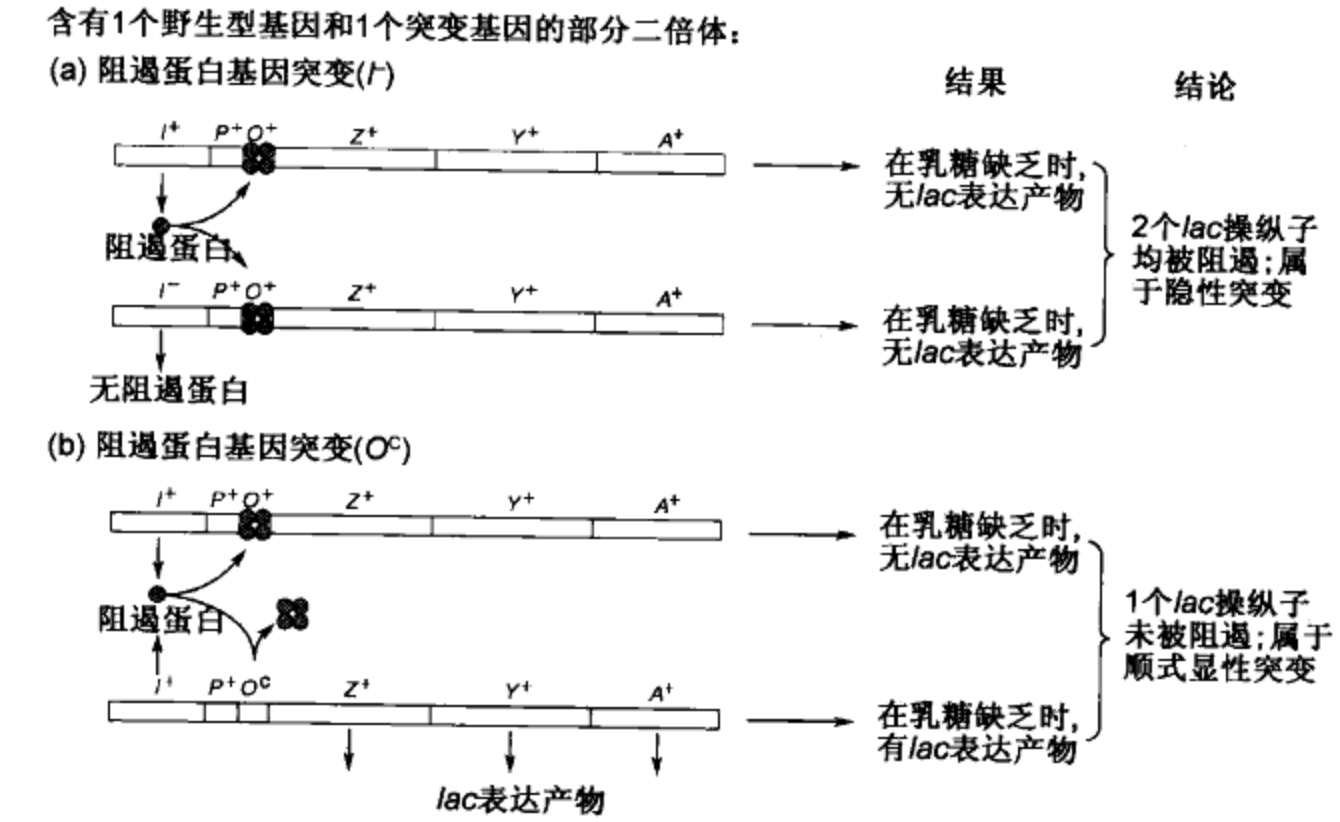


图 7.5　部分二倍体中 *lac* 操纵子调节突变的效应。Jacob、Monod 及其他研究者构建了 *E. coli* 部分二倍体菌株，如图（a）～（d）所示。然后在乳糖有或无条件下检测其 *lac* 产物。（a）部分二倍体有一个野生型操纵子（上）和一个阻遏基因突变（I^-）的操纵子（下）。野生型阻遏基因（I^+）可产生足够的正常阻遏物（绿色），阻遏 2 个操纵子，所以 I^- 突变是隐性的。（b）部分二倍体具有一个野生型操纵子（上）和一个操纵基因突变（O^c）的操纵子（下），突变操纵子不能与阻遏物（绿色）结合。这样，野生型操纵子是阻遏状态，而突变操纵子为去阻遏状态。无论乳糖是否存在，突变操纵子均可表达 *lac* 产物。由于只有与突变操纵基因连接的操纵子受到影响，该突变为顺式显性的。（c）部分二倍体具有一个野生型操纵子（上）和一个阻遏物基因突变（I^S）的操纵子（下），其产物（黄色）不能结合诱导物。因此突变的阻遏物不可逆地结合在两个操纵基因上，使 2 个操纵子都不可诱导，因此该突变为显性的。注意，这些阻遏物含有一些突变体亚基和野生型亚基，表现与突变蛋白一样。因此即使存在诱导物，也依然与操纵基因结合。（d）部分二倍体具有一个野生型操纵子（上）和一个阻遏物基因突变（I^{-d}）的操纵子（下），I^{-d}基因的产物（黄色）不能结合 *lac* 操纵基因。而且，由野生型和阻遏物单体形成的异源四聚体（heterotetramer）也不能与操纵基因结合。这样，即使在乳糖缺乏时，*lac* 操纵子仍然为去阻遏状态，所以该突变为显性的。此外，由于突变蛋白质破坏了野生型蛋白质的活性，所以称这种突变为显性负效应。

这两种普通类型的组成型突变（I^- 和 O^c）以相同方式影响 3 个 *lac* 基因（*lacZ*、*lacY* 和 *lacA*）的表达。遗传作图发现，这 3 个基因彼此相邻排列于染色体上，这些发现暗示操纵基因位于这 3 个结构基因的附近。

我们现在还了解了另一类组成型显性突变的阻遏物突变体（I^{-d}）。该突变基因编码的缺陷产物仍能与野生型阻遏物单体组合形成四聚体阻遏物。然而，缺陷单体破坏了四聚体的活性，妨碍了四聚体与操纵基因结合，因此使突变体呈现出相对于野生型的显性效应。在部分二倍体中，“损坏的”阻遏物与 2 个操纵基因都不能结合，这种突变不仅仅是顺式显性突变，这种“破坏性”突变在自然界中广泛存在，一般命名为**显性负效应**（dominant-negative）。

经过精湛的遗传学分析，Jacob 和 Monod 发展了操纵子概念，他们预测存在着 2 个关键调控元件：阻遏物基因和操纵基因。缺失突变分析表明第三个元件（启动子）也是调控 3 个 *lac* 基因表达所必需的。他们进一步推论这 3 个 *lac* 基因（*lacZ*、*lacY* 和 *lacA*）簇集在一起组成了一个调控单元，即乳糖操纵子。随后的生物化学研究充分证实了 Jacob 和 Monod 的完美假说。

小结　*lac* 操纵子的负调控机制如下：由于阻遏物蛋白阻止 RNA 聚合酶对 3 个 *lac* 基因的转录，所以当阻遏蛋白结合到操纵基因上时，操纵子即被关闭。当葡萄糖被消耗殆尽但有乳糖可被利用时，几个 *lac* 操纵子的酶分子（β-半乳糖苷酶）将少量乳糖转换为异乳糖。异乳糖作为诱导物与阻遏物结合，引起阻遏物构象改变，促使阻遏物与操纵基因解离。随着阻遏物的解离，RNA 聚合酶与启动子结合而起始 3 个 *lac* 基因的转录。遗传与生物化学实验研究结果显示，操纵基因和阻遏物基因是 *lac* 操纵子负调控的两个关键元件。

阻遏物与操纵基因间的相互作用

在 Jacob 和 Monod 的开创性研究工作之后，Walter Gilbert 和 Muller-Hill 等成功地部分纯化了 *lac* 阻遏物。这项工作的可贵之处在于它是在现代基因克隆技术产生之前的 20 世纪 60 年代完成的。Gilbert 和 Muller-Hill 的挑战是纯化细胞中含量甚微的蛋白质（*lac* 阻遏物），而又缺乏简便的方法鉴定分离的蛋白质。他们所采用的最灵敏的分析方法是将人工合成的标记诱导物**异丙基硫代半乳糖苷**（isopropylthiogalactoside，**IPTG**）与阻遏物结合。但是野生型细胞粗提取物中的阻遏物浓度太低，这种方法难以检测到阻遏物。为了克服这一难题，他们利用*E. coli*的阻遏蛋白基因突变菌株（$lacI^t$）进行阻遏蛋白的纯化，因为该突变株产生的阻遏蛋白与 IPTG 的结合更为紧密。这种紧密结合能力可使突变的阻遏蛋白结合足够的诱导物，因此即使在粗提物中也能检测到阻遏蛋白，这样 Gilbert 和 Muller-Hill 就可以纯化该蛋白质了。

Melvin Cohn 及同事将利用这一技术纯化得到阻遏物用于操纵基因结合研究。Cohn 及同事用硝酸纤维素滤膜结合实验（第 5、6 章）研究了阻遏物与操纵基因间的相互作用。如果阻遏物与操纵基因能正常相互作用，这种作用就能够被诱导物所阻断。确实，图 7.6 显示无诱导物时阻遏物-操纵基因结合的标准饱和曲线。但是存在合成诱导物 IPTG 时，二者不能结合。在另一个结合分析实验中（图 7.7），Cohn 小组发现与含野生型操纵基因的 DNA 相比，含组成型突变操纵基因（*lacO*^{c}）的 DNA 需要更高浓度的阻遏物才能实现二者的完全结合。这个实验为 Jacob 和 Monod 从遗传学上所定义的操纵基因是阻遏物结合位点提供了一个重要证据。否则，操纵基因突变是不会影响阻遏物的结合。

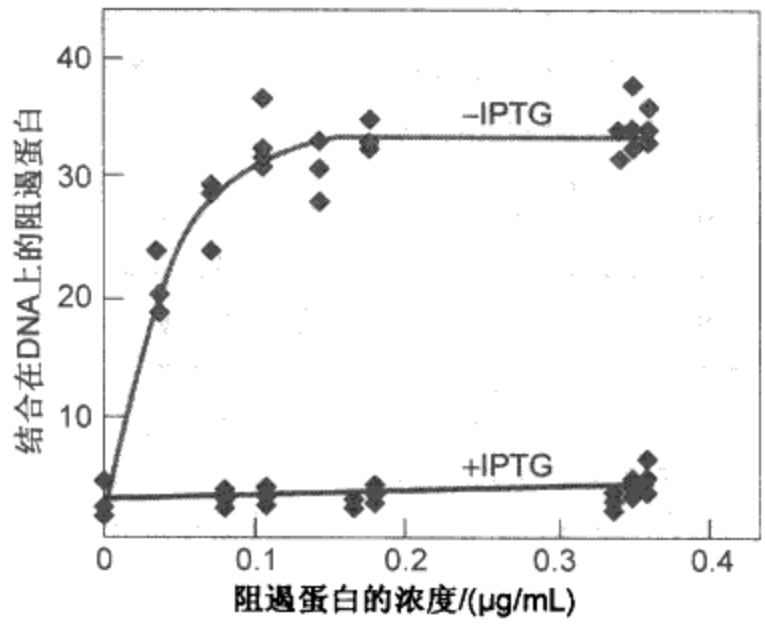


图 7.6　*lac* 操纵基因和阻遏物之间的结合分析。Cohn 及同事用^{32}P 标记含有 *lacO* 的 DNA，逐渐增加 *lac* 阻遏物的量，测定硝化纤维素膜上的放射性来分析阻遏物和操纵基因的结合。只有结合阻遏物的标记 DNA 才能黏附到硝酸纤维素膜上。－IPTG 表示无诱导物的阻遏物结合曲线，＋IPTG 表示有 1mmol/L IPTG 的阻遏物结合曲线，IPTG 阻止了阻遏物与操纵基因的结合。（Source：Adapted from Riggs，A.，et al.，On the assay，isolation，and characterization of the *lac* repressor，*Journal of Molecular Biology*，Vol. 34：366.）

小结　Cohn 及同事用滤膜结合法实验了 *lac* 阻遏物与操纵基因的结合。该实验还表明，与正常 *lac* 操纵基因相比，遗传学定义的组成型 *lac* 突变操纵基因对 *lac* 阻遏物的亲和力相对较低，这也表明通过遗传学方法和生物化学研究确定的操纵基因位点是同一个位点。

阻遏机制

尽管 Ira Pastan 及同事在 1971 年就已证明，即使

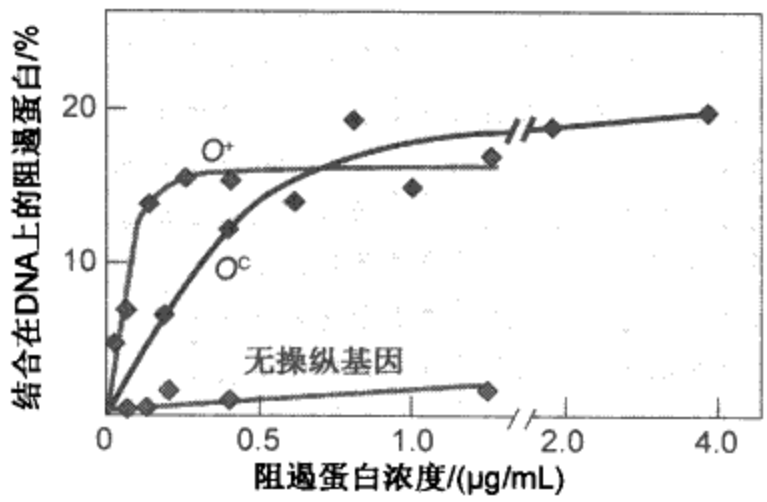


图 7.7　操纵基因 *lac O*^{c} 与阻遏物的亲和力低于野生型操纵纵基因与阻遏物的亲和力。Cohn 及同事按图 7.6 描述的方法进行 *lac* 操纵基因与阻遏物之间的结合力分析。选用了 3 种不同的 DNA 分子：DNA 分子中含有野生型操纵基因（O^+）；DNA 分子中含有组成型突变的操纵基因（O^c），对阻遏物的亲和力下降；以不有含 *lac* 操纵基因的 λΦ80 DNA 作对照。

有阻遏物存在，RNA 聚合酶也能与 *lac* 启动子紧密结合，但是多年来人们一直认为 *lac* 阻遏物是通过阻止 RNA 聚合酶结合启动子而发挥作用的。Pastan 的实验方案是在阻遏物存在条件下，将 RNA 聚合酶和含有乳糖操纵基因的 DNA 一起孵育，然后加入诱导物 IPTG 和利福平。本章后面将介绍，利福平将抑制转录，除非开放启动子复合体已形成（回顾第 6 章，开放启动子复合体是指 RNA 聚合酶在启动子处使 DNA 局部解链，并与之紧密结合）。在该实验中转录确实发生了，说明 *lac* 阻遏物没有阻止开放启动子复合体的形成，提示阻遏物没有阻止 RNA 聚合酶与 *lac* 启动子结合。Susan Straney 和 Donald Crothers 于 1987 年研究表明，RNA 聚合酶能够与阻遏蛋白共同结合在 *lac* 启动子上，进一步证实了 *lac* 阻遏蛋白不能阻止 RNA 聚合酶与 *lac* 启动子结合的观点。

如果我们接受这个观点，认为阻遏物与操纵基因结合后 RNA 聚合酶仍能紧密结合启动子，那么如何解释阻遏作用？Straney 和 Crothers 认为阻遏物阻止了开放复合体的形成，但这一观点又与 Pastan 观察到复合体具有利福平抗性的现象相矛盾。Barbara Krummel 和 Michael Chamberlin 提出了另一种不同的解释：阻遏蛋白能够阻止转录复合体从起始状态转换到延伸状态（第 6 章）。也就是说，阻遏物使 RNA 聚合酶处于无效转录状态，只能合成较短的寡聚核苷酸，无法完成启动子清除。

Jookyung Lee 和 Alex Goldfarb 为这一观点提供了实验证据。他们通过体外通读转录实验（第 5 章）证实，即使有阻遏物存在，RNA 聚合酶也已完成了与 DNA 模板的结合。具体的实验操作如下：首先，将阻遏物与 123bp 的 DNA 片段共孵育，该片段含有 *lac* 调控区和 *lacZ* 基因的起始序列，让阻遏物和操纵子结合 10min 后加入 RNA 聚合酶，之后再加入肝素［肝素是一种多聚阴离子（polyanion），能与游离的或松弛结合在 DNA 上的 RNA 聚合酶结合，使之失去与 DNA 结合的能力］。同时也加入了 RNA 聚合酶反应所需的其他成分，但不包括 CTP。最后，加入标记的 CTP 及 IPTG，并设不加 IPTG 的对照。实验的问题是：通读转录能否发生？如果能进行通读转录，那么 RNA 聚合酶就能够与启动子形成对肝素有抗性的（开放）复合体，即使是在阻遏蛋白存在的情况下。事实上，如图 7.8 所示，好像

图 7.8　在体外，即使有 *lac* 阻遏物，RNA 聚合酶也能与 *Lac* 启动子形成开放启动子复合体。Lee 和 Goldfarb 将含有 *lac* UV5 启动子的 DNA 片段单独温育（泳道 1）或与 *lac* 阻遏蛋白共同温育（泳道 2 和 3），当阻遏蛋白与操纵基因完成结合后，研究者向反应体系中添加 RNA 聚合酶，温育 20min 以便开放启动子复合体的形成，然后添加肝素以阻止开放启动子复合体的进一步形成，同时也添加除 CTP 以外的其他反应组分，温育 5min 后，向反应体系中单独添加［α-^{32}P］CTP 或同时添加诱导物 IPTG，温育 10min 进行 RNA 的合成，然后对合成的转录本进行凝胶电泳。泳道 3 的结果表明，即使阻遏蛋白先于 RNA 聚合酶与 DNA 结合，仍然能够发生转录事件。由此可见，阻遏蛋白并不能阻止 RNA 聚合酶与启动子结合而形成开放启动子复合体。（*Source*: Lee J. , and Goldfarb A. , *lac* repressor acts by modifying the initial transcribing complex so that it cannot leave the promoter. *Cell* 66(23 Aug 1991) f. 1, p. 794. Reprinted by permission of Elsevier Science.）

阻遏蛋白根本就不存在一样，确实发生了通读转录。由此可见，在这些体外条件下，阻遏蛋白似乎并不能阻止 RNA 聚合酶与 *lac* 启动子间的紧密结合。

如果不通过阻止 RNA 聚合酶与启动子结合而阻断 *lac* 操纵子的转录，那么 *lac* 阻遏物如何发挥自身的阻遏功能？Lee 和 Goldfarb 注意到，在阻遏物存在时，流产性转录（第 6 章）的长度仅约为 6nt，而无阻遏物时，流产性转录物的长度约为 9nt。在阻遏物存在条件下能够形成转录本（即使很短）的事实有力地支持了以下结论：至少在体外实验条件下，当存在阻遏物时，RNA 聚合酶确实能够与 *lac* 启动子结合。该实验结果也提示，阻遏物能够将 RNA 聚合酶限定在只能合成流产性转录物的无效生产状态，从而阻断了 *lac* 操纵子的转录。这样 RNA 聚合酶就不可能进入转录延伸阶段。

在 Lee 和 Goldfarb 以及其他研究者采用的研究方案中，存在着一个问题即实验是在离体非生理状态下进行的。例如，蛋白质（RNA 聚合酶、阻遏物）浓度要比体内的浓度高。为解决这个问题，Thomas Record 及同事在尽量是体内外条件一致的基础上开展了离体动力学研究。先于体外使 RNA 聚合酶与 *lac* 启动子形成复合体，然后测定复合体独自合成流产性转录物的速率，或者在添加肝素或 *lac* 阻遏物的情况下合成流产性的转录物的速率。为测定转录速率，研究者在反应体系中添加了 UTP 的同系物 *pppU，即 Y-磷酸具有一个荧光标签的 UTP，当 UMP 掺入到 RNA 链中后，*pppU 就释放出带有标签的焦磷酸（*pp），使体系中的荧光强度增加。图 7.9 的结果表明，在无竞争者存在的情况下，流产性转录物的合成维持在一个较高的水平上，但在肝素或阻遏物存在时，转录的水平就会迅速下降。

Record 及同事对这些结果的解释如下：聚合酶-启动子复合物与游离的聚合酶、启动子三者处于平衡状态，竞争者不存在时（曲线 1），解离的聚合酶能够重新与启动子结合，继续进行流产性转录物的合成。然而，肝素（曲线 2）和阻遏蛋白（曲线 3）可阻止 RNA 聚合酶与启动子的重新结合。肝素通过与聚合酶结合而阻止了聚合酶与 DNA 的结合；阻遏蛋白可能是通过与邻近启动子的操纵基因结合而阻止了 RNA 聚合酶与启动子的结合。由此可见，这些研究数据支持了过去所认为的 RNA 聚合酶与阻遏蛋白之间存在竞争的假说。

对 *lac* 阻遏物作用机制的研究迂回曲折，最新研究结果表明最初的竞争假说是正确的，但这还不是阻遏机制研究的尾声。

lac 操纵子阻遏机制的另一个因素是存在 3 个操纵基因而不是 1 个，即位于转录起始位点附近的主操

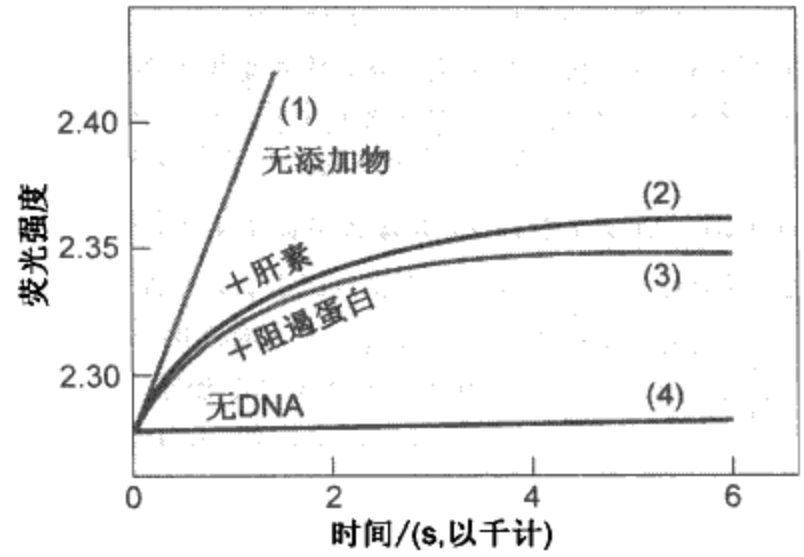


图 7.9　*lac* 阻遏物对 RNA 聚合酶与 lac 启动子解离的影响。Record 及同事将 RNA 聚合酶与含有 *lac* 启动子和操纵基因序列的 DNA 片段混合，形成 RNA 聚合酶-*lac* 启动子复合体。在 γ-磷带有荧光标签的 UTP 类似物 *pppU 存在情况下，让复合体转录 DNA 片段合成流产性转录物。当 UMP 掺入 RNA 链时，释放标记的焦磷酸（*pp）使荧光强度增加。实验分 3 组：无添加物（曲线 1）；加入肝素，与从 DNA 解离的 RNA 聚合酶结合而阻止转录重新起始（曲线 2）；添加低浓度 *lac* 阻遏物（曲线 3）；无 DNA 实验为对照（曲线 4）。同肝素一样，阻遏物也能抑制中断转录的重新起始，表明阻遏物可阻断解离 RNA 聚合酶与启动子的重新结合。（*Source*: Adapted from Schlax, P. J., Capp, M. W., and M. T. Record, Jr. Inhibition of transcription initiation by *lac* repressor, *Journal of Molecular Biology* 245:331-50）

纵基因（major operator）和 2 个辅操纵基因（auxiliary operator）分别位于转录起始位点的上游和下游。图 7.10 给出了这 3 个操纵基因的空间排列，典型（主）操纵基因 O_1 以 +11 为序列中心，下游辅操纵基因 O_2 以 +412 为序列中心，上游辅操纵基因 O_3 以 −82 为序列中心。我们已讨论过主操纵基因及其作

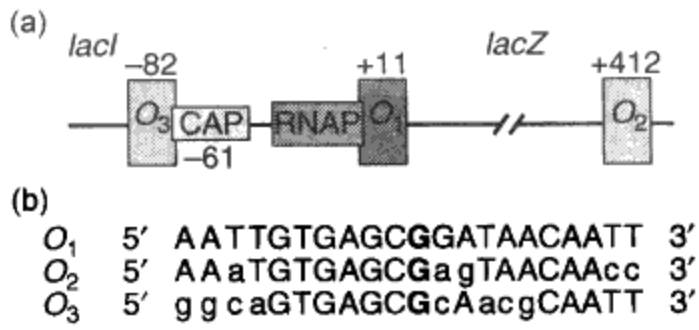


图 7.10　3 个 *lac* 操纵基因。（a）*lac* 调控区的遗传图谱。主操纵基因（O_1）为红色；2 个辅操纵基因为紫色；CAP 结合位点和 RNA 聚合酶结合位点分别为黄色和蓝色。CAP 是 *lac* 操纵子的正调控因子，本章下一节再做讨论。（b）3 个操纵基因的序列。3 个操纵基因的序列比对，粗体 G 为序列的中心，在 2 个辅操纵基因（O_2 和 O_3）序列中的小写字母表示与主操纵基因不同的碱基。

用，研究者一般认为主操纵基因单独发挥作用。Benno Müller-Hill 和其他研究者近来对辅操纵基因进行了较为详尽的研究，发现辅操纵基因并不是主操纵基因无关紧要的拷贝，它们在阻抑过程中起重要作用。Müller-Hill 及同事通过研究证实了辅操纵基因的作用，如果除去其中的任何一个辅操纵基因，只会轻微降低阻遏效率。但是，如果同时去除 2 个辅操纵基因，则会使阻遏效率降为原先的 1/50。图 7.11 显示了这些实验结果：3 个操纵基因同时存在时，抑制转录的效率提高了 1300 倍；只有 2 操纵基因存在时会使抑制转录的效率提高 400～700 倍；而主操纵基因自身抑制转录的效率仅为 18 倍。

1996 年，Mitchell Lewis 及同事为 3 个操纵基因间的协同关系提供了结构基础。他们检测了 *lac* 阻遏物与包含操纵基因序列的 21bp DNA 片段形成复合体的晶体结构。从图 7.12 可以看出，四聚体阻遏物中 2 个二聚体都是相互独立的 DNA 结合实体，能够与 DNA 大沟相互作用。也可以清楚地看出四聚体中两个二聚体分别与两个不同的操纵基因序列相结合，这就很容易理解为什么这两个操纵基因是同一条长 DNA 片段中的一部分。

小结 有两种竞争假说解释了 *lac* 操纵子的抑制机制。一种认为，阻遏物存在时 RNA 聚合酶与 *lac* 启动子结合，但阻遏物抑制了 RNA 聚合酶由流产性转录向进行性转录的过渡；另一种认

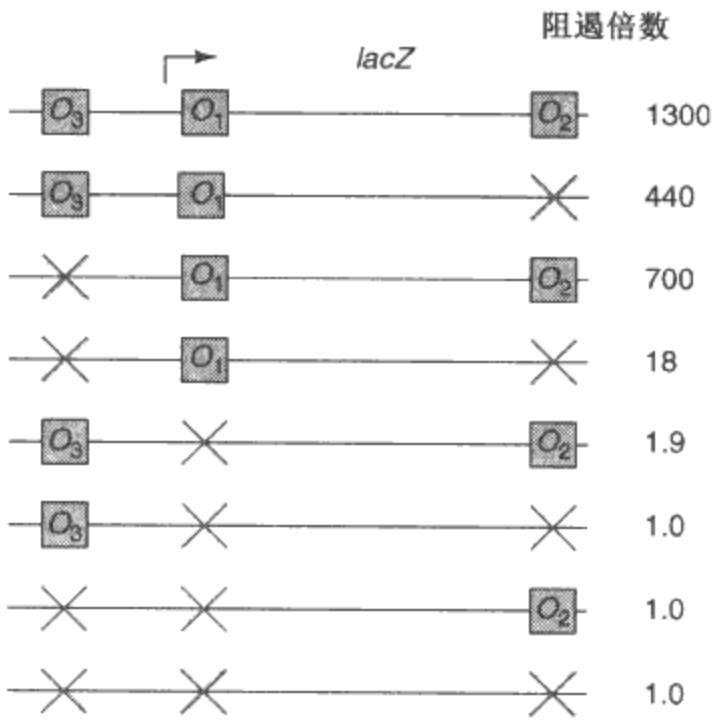


图 7.11 3 个 *lac* 操纵基因的突变效应。Müller-Hill 及同事将野生型和突变型 *lac* 操纵子片段插入 λ 噬菌体 DNA，然后侵染 *E. coli* 细胞，使 λ 噬菌体处于溶源周期（第 8 章）。将包含 3 个操纵基因和 *lacZ* 基因的 *lac* 操纵子片段整合到细菌基因组中，细菌基因组不含有其他 *lacZ* 基因，但拥有野生型 *lacI* 基因。分析 β-半乳糖苷酶在诱导物 IPTG 有或无条件下的合成情况，根据酶的活性推知阻遏效率。图右侧为两种条件下阻遏活性的比率。当 3 个操纵基因都存在时，有诱导物时酶活性是无诱导物时酶活性的 1300 倍，即阻遏效率达 1300 倍；λEwt123（顶部）携带 3 个野生型操纵基因（绿色），其他噬菌体缺失 1 个或多个操纵基因（红色×）。（*Source*：Adapted from Oehler，S.，E. R. Eismann，H. Krämer，and B. Müller-Hill. 1990. The three operators of the *lac* operon cooperate in repression. *The EMBO Journal* 9：973-79.）

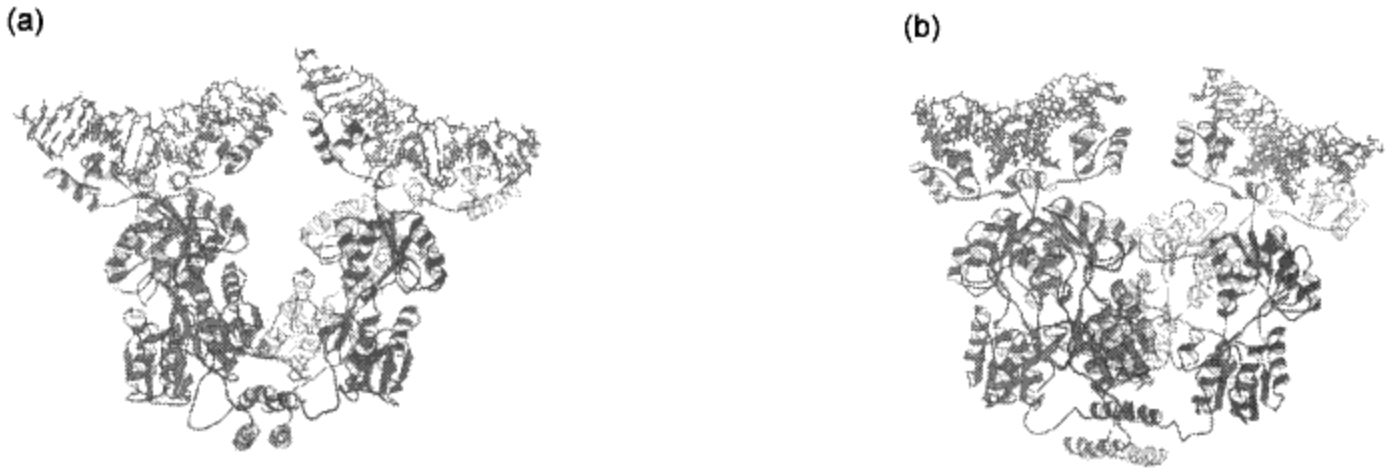


图 7.12 结合有两个操纵基因片段的 *lac* 阻遏物四聚体结构。Lewis、Lu 及同事对结合 DNA 的 *lac* 阻遏物进行 X 射线衍射晶体学分析，长 21bp 的 DNA 片段包含操纵基因序列。4 个阻遏物单体分别用紫色、绿色、黄色、红色表示。2 个阻遏物二聚体在底部相互作用形成四聚体，每个二聚体拥有 2 个 DNA 结合域，从结构图顶部可以看出二聚体与 DNA 大沟相互作用。该结构清楚地显示 2 个二聚体分别独立地与不同 *lac* 操纵基因结合。（a）和（b）为同一结构的前视图和侧视图。（*Source*：Lewis et al.，Crystal structure of the lactose operon processor and its complexes with DNA and inducer. *Science* 271（1 Mar 1996），f. 6，p. 125. © AAAS.）

为，阻遏物与操纵基因相结合，阻碍 RNA 聚合酶与邻近操纵基因的启动子结合。最近的研究支持后者。除了与启动子相邻的典型（主）操纵基因外，还存在 2 个辅操纵基因，分别位于转录起始位点的上游和下游。这 3 个操纵基因都是高效阻遏所必需的，两个操纵基因同时参与阻遏是比较理想的，主操纵基因自身只能引发中等水平的阻遏效应。

lac 操纵子的正调控

在本章前面我们已经了解到只要有葡萄糖存在，*E. coli*细胞就能够保持乳糖操纵子处于相对失活状态。*E. coli*细胞优先利用葡萄糖作为能源，而不用其他碳源，这种选择是由葡萄糖的降解产物或代谢产物（catabolite）控制的，所以将这种调控方式称为**代谢阻遏**（catabolite repression）。

理想的 *lac* 操纵子正调控因子应该能感知葡萄糖缺乏，并通过激活 *lac* 启动子做出应答，使 RNA 聚合酶与启动子结合而转录结构基因（当然，前提条件是假定有乳糖存在，阻遏物因而不能与操纵基因结合）。能够对葡萄糖浓度变化做出应答的物质是**环化腺苷酸（cyclic-AMP，cAMP）**（图 7.13）。当葡萄糖浓度降低时，cAMP 浓度增加。

代谢物激活蛋白 Ira Pastan 及同事通过研究表明，cAMP 能够克服 *lac* 以及 *gal*、*ara* 等操纵子的代谢物阻遏效应。*gal* 和 *ara* 操纵子分别控制半乳糖和阿拉伯糖的代谢。也就是说，即使在有葡萄糖存在的情况下，cAMP 也能使操纵子中的结构基因活化。这一发现暗示 cAMP 对 *lac* 操纵子具有强烈的正调控效应。这是否意味着 cAMP 就是正效应物？但这并不准确，因为 *lac* 操纵子的正调控因子实质是一个复合体，由 cAMP 和蛋白质因子两部分构成。

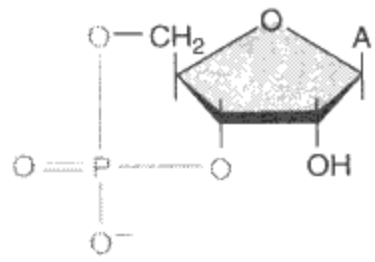


图 7.13　环化腺苷酸 cAMP。注意图中的 5′→3′磷酸二酯键（蓝色）。

环腺苷酸

Ceoffrey Zubay 及同事研究发现，在提供 cAMP 的条件下，*E. coli* 的无细胞粗提物能够合成 β-半乳糖苷酶，这一发现为从 *E. coli* 的无细胞粗提物中获得 cAMP 激活结构基因表达所必需的蛋白质提供了思路。Zubay 将这种蛋白质称为代谢激活蛋白（catabolite activator protein，CAP），后来，Pastan 研究小组也发现了这种蛋白质并命名为环腺苷酸受体蛋白（cyclic-AMP receptor protein，CRP）。为避免混淆，在以后所讨论的研究实验中均将这种蛋白质称为 CAP，但是编码该蛋白质的基因已被正式命名为 *crp*。

Pastan 测定出 CAP-cAMP 复合物的解离常数为 $1\times10^{-6}\sim1\times10^{-6}$ mol/L。Pastan 发现 CAP-cAMP 复合体的解离常数为 $1\sim2\times10^{-6}$ mol/L，但研究者也分离获得了一株突变体，该突变体的 CAP 与 cAMP 间的结合力大约为原先的 1/10，如果 CAP-cAMP 确实是 *lac* 操纵子正调控的重要因子，那么，向突变体的无细胞提取物添加 cAMP 后，其 β-半乳糖苷酶的合成量就应低于野生型无细胞提取物的 β-半乳糖苷酶合成量。图 7.14 显示实验结果的确如此。为使这一结论更加可信，Pastan 证明突变株提取物（加 cAMP）加入野生型 CAP 后，β-半乳糖苷酶的合成率提高约 3 倍。

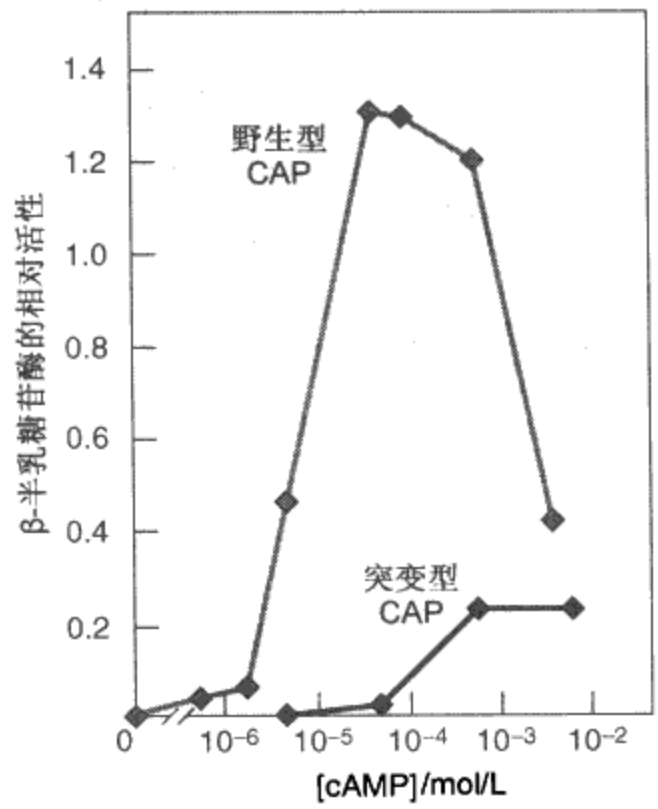


图 7.14　cAMP 与野生型和突变型 CAP 均能促进 β-半乳糖苷酶的合成。Pastan 及同事逐渐增加 cAMP 浓度，激活野生型和突变体无细胞提取物生产 β-半乳糖苷酶。红色表示野生型，绿色表示突变体（CAP 对 cAMP 的亲和力降低了）。突变体产生很少的 β-半乳糖苷酶，如果 CAP-cAMP 复合物在 *lac* 操纵子转录中具有重要作用，那么这一结果正是我们所期望的。在野生型细胞提取物中，太多 cAMP 明显干扰了 β-半乳糖苷酶的合成，这不奇怪，因为 cAMP 有很多作用，一些 cAMP 可能间接抑制了 *lacZ* 基因体外表达的某些步骤。（*Source*: Adapted from Emmer, M., et al., Cyclic AMP receptor protein of *E. coli*: Its role in the synthesis of inducible enzymes, *Proceedings of the National Academy of Sciences* 66 (2): 480-487, June 1970.）

小结 *lac* 操纵子及其他几种编码糖代谢酶的可诱导型操纵子，其正调控由代谢物激活蛋白（CAP）所介导，该蛋白和 cAMP 共同激活转录。葡萄糖能够降低 cAMP 的浓度，抑制转录的激活。只有当葡萄糖浓度很低，且需要代谢另一种能源时，*lac* 操纵子才被激活。

CAP 的作用机制

CAP 和 cAMP 如何促进 *lac* 操纵子转录？Zubay 及同事发现了一类 *lac* 突变体，其 CAP 和 cAMP 不能促进 *lac* 操纵子转录。遗传作图发现突变位于 *lac* 启动子上，表明 CAP-cAMP 复合体的结合位点在 *lac* 启动子内，随后的分子生物学研究显示，CAP-cAMP 的结合位点［**激活因子结合位点** (activator-binding site)］位于启动子的上游。Pastan 及同事进一步研究表明 CAP-cAMP 与激活因子位点的结合有助于 RNA 聚合酶在启动子处形成开放启动子复合体。

图 7.15 描述了实验过程。首先，在 CAP-cAMP 有或无两种情况下使 RNA 聚合酶与 *lac* 启动子结合，然后向反应体系中同时添加核苷酸和利福平，分析开放启动子复合体是否形成。如果在利福平加入之前无开放启动子复合体形成，则利福平抑制了转录起始，因为 DNA 熔解耗时长，其间足以使抗生素在转录起始前抑制 RNA 聚合酶的活性。然而，如果在利福平加入前开放启动子复合体已形成，那么该复合体就能够进行核苷酸的聚合，因为核苷酸先于抗生素与 RNA 聚合酶结合，使 RNA 聚合酶能够起始转录。一旦起始了 RNA 链的合成，RNA 聚合酶就会获得利福平抗性，直至完成 RNA 链的合成。事实上，Pastan 及同事发现，在缺乏 CAP-cAMP 情况下形成的聚合酶-启动子复合体对利福平敏感，说明未形成开放启动子复合体。而 CAP-cAMP 存在时形成利福平抗性开放启动子复合体。

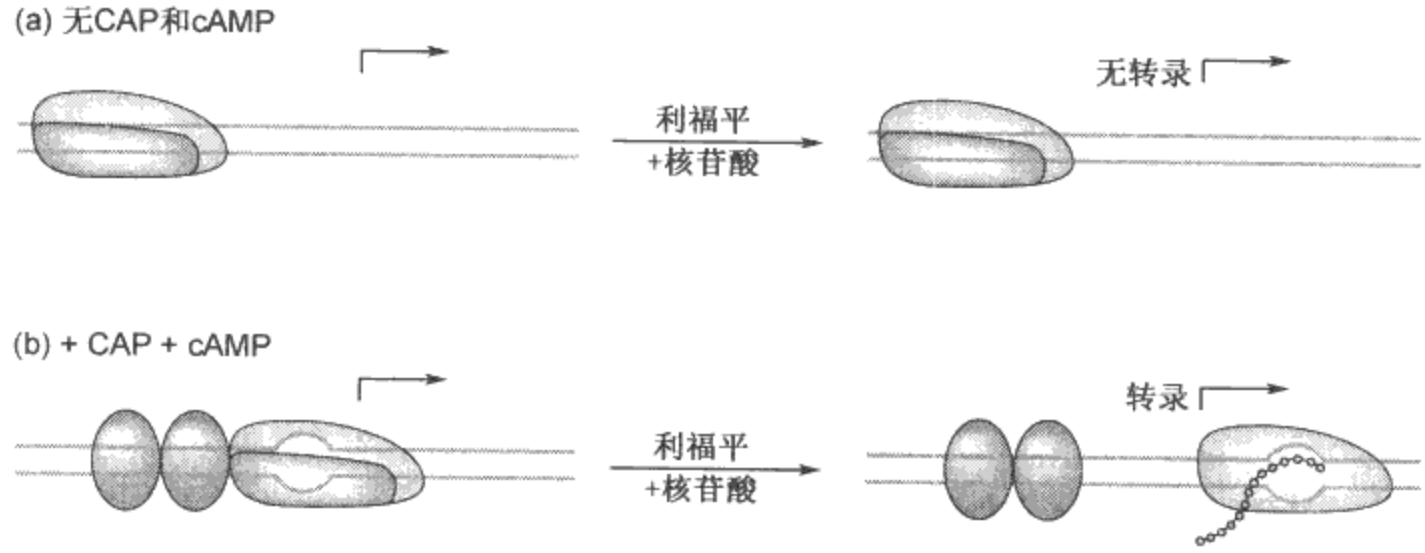


图 7.15 CAP 与 cAMP 协同促进开放启动子复合体的形成。（a）无 CAP 时，RNA 聚合酶与含 *lac* 启动子 DNA 片段随机松散地结合，在同时加入利福平及核苷酸后，这种结合受利福平抑制，因此无转录发生；（b）有 CAP 和 cAMP（紫色）存在时，RNA 聚合酶与 *lac* 启动子结合形成开放启动子复合体，在同时加入利福平及核苷酸后，这种结合不受利福平抑制，因为核苷酸先于利福平与聚合酶结合，使开放启动子复合体能够启动核苷酸聚合。一旦第一个磷酸二酯键形成，RNA 聚合酶在重新起始转录前一直具有利福平抗性。在这些条件下能够转录，表明 CAP 和 cAMP 促进开放启动子复合体形成。绿色线条代表 RNA 分子。

图 7.15b 显示，CAP-cAMP 二聚体结合在左侧的激活因子位点上，聚合酶结合在右侧的启动子上。我们如何知道这就是 CAP-cAMP 及 RNA 聚合酶在 DNA 上的正确结合顺序呢？最初证据来自遗传学实验。启动子的左侧序列突变，可阻止 CAP-cAMP 对转录的促进作用，但仍存在低水平转录。如图 7.16 所示，缺失突变体 L1 完全消除了 CAP 和 cAMP 对 *lac* 操纵子的正调控作用，说明 CAP 结合位点至少部分位于删除区域内。而 L1 的缺失对 CAP 非依赖性转录没有影响，所以删除区域不包括与 RNA 聚合酶结合的启动子部分。因此，缺失序列的右末端大体上就是激活因子结合位点与启动子之间的分界线。

lac、*gal* 和 *ara* 操纵子的 CAP 结合位点均含有 TGTGA 序列，保守性表明 TGTGA 序列是 CAP 结合位点的重要组成部分，目前已有直接证据证明该序列的重要性。例如，DNA 足迹研究表明，CAP-cAMP 复合体的结合能够保护该序列中的 G 免受硫酸二甲酯的甲基化修饰，说明 CAP-cAMP 复合体与 G 紧密结合，使 G 免受甲基化试剂的攻击。

lac 操纵子及其他能受 CAP 和 cAMP 激活的操纵子都具有非常弱的启动子，其－35 框与保守序列

不同，很难识别。其实这种情形并不足为奇，因为，如果 *lac* 操纵子拥有一个很强的启动子，那么无需 CAP 和 cAMP 帮助，RNA 聚合酶自身就能结合启动子而形成开放启动子复合体，这样即使存在葡萄糖，*lac* 操纵子也处于激活状态。所以该启动子必须是依赖于 CAP 和 cAMP 的弱启动子。事实上，已发现存在不依赖于 CAP 和 cAMP 的 *lac* 强突变启动子（如 *lac* UV5 启动子）。

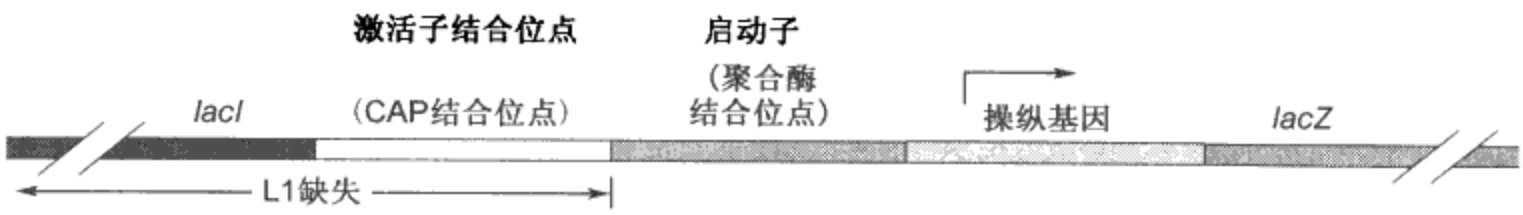


图 7.16 *lac* 调控区。紧邻操纵基因上游的激活子-启动子区域包含有位于左侧的激活子结合位点（CAP 结合位点）（黄色）、位于右侧的启动子（聚合酶结合位点）（紫色）。这些位点是通过 DNA 足迹实验和遗传学分析而确定的。L1 缺失突变就是遗传学分析的一个事例，L1 突变体的 *lac* 操纵子存在本底水平的转录，但不会被 CAP 和 cAMP 所促进，说明该突变体具有完整的启动子序例，但激活子结合位点发生了缺失。

小结 CAP-cAMP 复合体通过结合在邻近启动子的激活因子结合位点上而促进 *lac* 操纵子的转录，这种结合也有助于 RNA 聚合酶与启动子的结合。

募集 CAP-cAMP 复合体如何将 RNA 聚合酶募集到启动子上？这种募集作用（recruitment）包括两步：①形成闭合启动子复合体；②将闭合启动子复合体转换为开放启动子复合体。William McClure 及同事将这两步过程归纳为如下反应式：

$$R+P \underset{K_B}{\rightleftharpoons} RP_C \xrightarrow[k_2]{} RP_O$$

式中，R 代表 RNA 聚合酶；P 代表启动子；RP_C 代表闭合启动子复合体；RP_O 代表开放启动子复合体。McClure 及同事利用动力学方法来区分这两步反应，确定 CAP-cAMP 通过提高 K_B 而直接促进第一步反应，但对 k_2 几乎没有影响，所以不会促进第二步反应。虽然如此，通过提高闭合启动子复合体的形成速率，CAP-cAMP 为开放启动子复合体的转换提供了更多原料（即闭合启动子复合体）。CAP-cAMP 的净效应是增加开放启动子复合体的形成速率。

Philip Malan 和 McClure 也证明 CAP-cAMP 还有另一种方式促进乳糖操纵子转录。William Reznikoff 及同事在 1978 年就已证实乳糖操纵子在主启动子上游 22bp 处有另一个可选择性启动子。McClure 将主启动子命名为 P_1，可选择性启动子命名为 P_2。Malan 和 McClure 的体外转录实验表明，CAP-cAMP 降低 P_2 处的转录起始，而促进 P_1 处的转录起始。因为 P_2 似乎是个无效启动子，能够限定聚合酶结合的数量，使更多的游离聚合酶与更有效的启动子 P_1 结合，从而提高 *lac* 操纵子的转录，这就是 CAP-cAMP 可能的作用机制。

结合在激活子结合位点上的 CAP-cAMP 如何辅助 RNA 聚合酶与启动子的结合？一种长期被支持的假说认为，当 CAP 和 RNA 聚合酶与各自的 DNA 靶位点结合后，两者会发生直接的接触，因此它们与 DNA 的结合具有协同效应。

这一假说获得了较多实验证据的支持。第一，在有 cAMP 时，通过超速离心可使 CAP 和 RNA 聚合酶共沉淀，表明两者之间具有亲和性。第二，当 CAP 和 RNA 聚合酶与各自的 DNA 靶位点结合后，彼此能发生化学交联，表明两者彼此十分靠近。第三，DNase 足迹实验（第 5 章）显示 CAP-cAMP 的足迹与聚合酶的足迹相邻。因此两者的 DNA 结合位点距离很近，足以使它们在各自的 DNA 位点上结合后能够发生相互作用。第四，某些 CAP 突变会降低转录激活作用，但不影响与 DNA 的结合（或弯曲）。其中一些突变改变了与聚合酶相互作用的 CAP 区域的氨基酸［**激活区域** activation region I（AR I）］。第五，推测与 CAP 的 AR I 相互作用的位点是聚合酶 α 亚基的羧基端结构域（αCTD），缺失 αCTD 可阻止 CAP-cAMP 介导的激活作用。第六，Richard Ebright 及同事在 2002 年对 DNA、CAP-cAMP 和 RNA 聚合酶 αCTD 复合体进行 X 射线晶体结构分析，结果显示尽管 CAP 与 RNA 聚合酶的界面不大，但 AR I 确实与 αCTD 相接触。为了使 αCTD 自己结合到复合体上，他们将 CAP 结合位点两侧的碱基序列变成富含 A-T 序列（5′-AAAAAA-3′），以便吸引 αCTD 结合。图 7.17a 是所测定的晶体结构，1 分子 αCTD（αCTD^{DNA}）与 DNA 结合，另 1 分子 αCTD（$\alpha CTD^{CAP\cdot DNA}$）与 DNA 和 CAP 都结合。后一种情况的 αCTD 很明显与 CAP 的 AR I 接触，详细的结构分析准确地确定了每个蛋白质中参与相互作用的氨基酸残基。只有一个 αCTD 单体与一个 CAP 单体结合是体内真实情形的反应，另一个 αCTD 单体不与 CAP 接触或者是晶体结构的特征，或者是体内情况就是如此。

图 7.17　CAP-cAMP-αCTD-DNA 复合体及 CAP-cAMP-DNA 复合体的晶体结构。（a）CAP-cAMP-αCTD-DNA 复合体。DNA 为红色，CAP 为青色，cAMP 用细红线表示，αCTD^{DNA} 为深绿色，$\alpha CTD^{CAP,\ DNA}$ 为浅绿色；（b）CAP-cAMP-DNA 复合体。各部分的颜色同图（a）。（*Source*：Benoff et al.，*Science* 297 © 2002 by the AAAS.）

在图 7.17a 中另一点值得关注的是当 CAP-cAMP 与 DNA 靶位点结合后，引起 DNA 片段产生约 100°的弯曲。早在 1991 年，Thomas Steitz 及同事在缺少 αCTD 的 CAP-cAMP-DNA 复合体晶体结构图像中就注意到了 DNA 弯曲现象。在 Ebright 及同事获得 CAP-cAMP-DNA 复合体的晶体结构图像中同样能观察到 DNA 弯曲（图 7.17b）。在 CAP-cAMP-αCTD-DNA 复合体和 CAP-cAMP-DNA 复合体中，DNA 及 CAP 的结构完全相同，说明 αCTD 不会扰乱复合体的结构。

早在 1984 年，Hen-Ming Wu 和 Donald Crothers 就通过**电泳检测 DNA 弯曲**（detecting DNA bending by electrophoresis）（图 7.18）。当 DNA 片段弯曲后，其电泳迁移率较慢，而且弯曲部位越接近 DNA 片段中部，DNA 分子的迁移率就越慢（图 7.18b、c）。Wu 和 Crothers 利用不同构型的 DNA 分子电泳迁移率不同的特点，制备了 *lac* 操纵子的 DNA 片段，各片段的长度相同，但 CAP 结合位点不同。然后使每一 DNA 片段分别与 CAP-cAMP 结合，对形成的 DNA-蛋白质复合体进行凝胶电泳。如果 CAP 的结合确实导致 DNA 弯曲，那么不同的 DNA 片段就会有不同的电泳迁移率。如果 CAP 的结合不会使 DNA 弯曲，则所有 DNA 片段具有相同电泳迁移率。图 7.19d 表明不同 DNA 片段确实具有不同的迁移率，而且 DNA 弯曲越显著，电泳迁移率相差越大。换句话说，根据图 7.18 提供的曲线，可以估算出 CAP-cAMP 结合后引起 DNA 弯曲的程度。CAP-cAMP 结合可使 DNA 发生约 90°的弯曲，与 X 射线衍射晶体学研究得出的 100°弯曲相吻合。这种弯曲可能是复合体中 DNA 与蛋白质发生最佳相互作用所必需的。

我们引用的所有研究均指出了 CAP 与 RNA 聚合酶，特别是聚合酶的 αCTD 间发生蛋白质-蛋白质相互作用的重要性。该假说推测 αCTD 缺失突变将会阻止 CAP-cAMP 对转录的激活。Kazuhiko Igarashi 和 Akira Ishihama 为 RNA 聚合酶 αCTD 在 CAP-cAMP 激活转录中的重要性提供了遗传学证据。研究者首先将各个分离的亚基在体外重组为功能性 RNA 聚合酶，各亚基均为野生型，除了有些实验中 α 亚基为缺失 CTD 的突变型。一种突变型 α 亚基缺失第 256 位氨基酸残基之后的碳末端序列（正常 α 亚基含有 329 个氨基酸）。另一种突变型 α 亚基缺失第 235 位氨基酸残基之后的碳末端序列，然后利用这个重组的 RNA 聚合酶进行体外通读转录。表 7.2 为体外通读转录的实验结果，在有或无 CAP-cAMP 时用野生型重组酶或突变型 α 亚基的重组酶在 CAP-cAMP 依赖性 *lac* 启动子（P1）或非依赖性 *lac* 启动子（*lac* UV5）处起始转录。正如推测的那样，CAP-cAMP 不能促进 *lac* UV5 启动子的转录，因为 *lac* UV5 是一个强启动子，对 CAP-cAMP 不敏感。CAP-cAMP 能够使 *lac*P1 启动子的转录效率提高 14 倍。有趣的是携带截短型 α 亚基的重组酶在无 CAP-cAMP 时，从任一启动子处起始的转录与野生型聚合酶没有区别，只是不能被 CAP-cAMP 所促进。由此可见，αCTD 并不是重组功能性 RNA 聚合酶所必需的，即 αCTD 的缺失不会影响重组酶的活性，但 αCTD 却是 CAP-cAMP 发挥促进转录功能所必需的。

图 7.19 概述了激活假说的要点。CAP-cAMP 二聚体在与其激活因子位点结合的同时，也与聚合酶 α 亚基的 C 端（αCTD）发生结合，促进了 RNA 聚合酶与启动子的结合。这种结合可能同 αCTD 与 DNA 的 UP 元件的结合（第 6 章）有功能对等效应，因而能够增强聚合酶与启动子的结合。

可促进 100 多个启动子起始转录的 CAP 只是细菌转录激活因子中的一员，现在越来越多的转录激活子被发现。在第 9 章将举例详细介绍。

小结　CAP-cAMP 与激活因子结合位点结合后，招募 RNA 聚合酶与 *lac* 启动子结合，形成闭

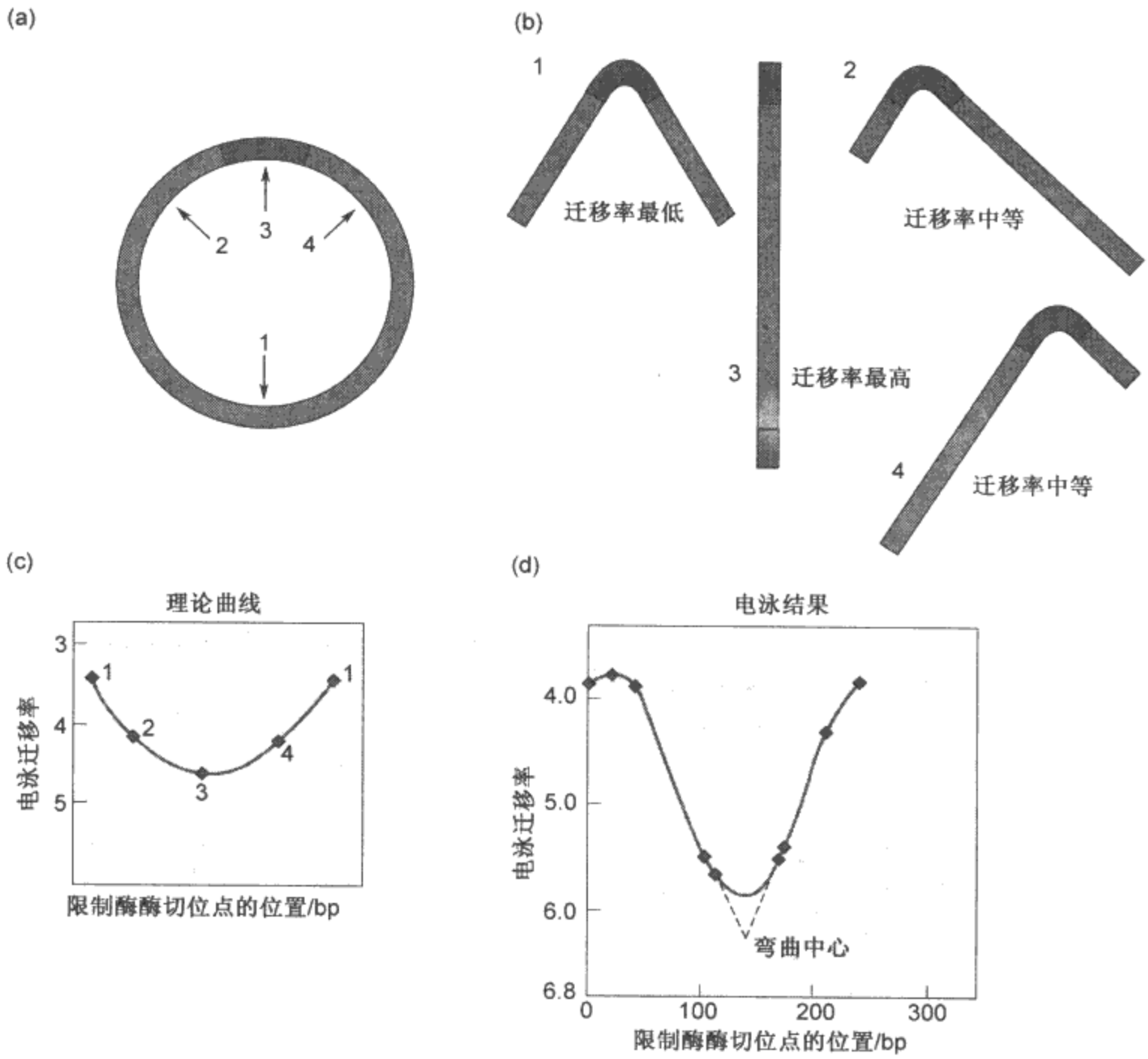


图 7.18　CAP-cAMP-启动子复合体的电泳特征。（a）假定的环形 DNA 图谱。其中心为蛋白质结合位点（红色），并具有 4 个不同的限制性内切核酸酶（简称限制酶）的酶切位点（箭头所示）。（b）将图（a）所示的 DNA 分子用限制酶进行酶切，每个 DNA 分子只能用一种酶进行酶切。然后加入 DNA 结合蛋白，使 DNA 弯曲。限制酶 1 在结合位点的对面将 DNA 分子切割，使蛋白质结合位点处于线性 DNA 的中间。限制酶 2 和 4 切割后，使蛋白质结合位点远离将 DNA 分子中心。限制酶 3 在结合位点内切割，几乎不引起 DNA 弯曲，即使有弯曲也是十分轻微的弯曲。（c）电泳迁移率与弯曲发生在不同位点的 DNA 的理论曲线。中部弯曲的 DNA 迁移率最低（曲线的两端），沿 *y* 轴向下迁移率逐渐增加。（d）CAP-cAMP 与 *lac* 启动子在不同位点的 DNA 片段结合后的电泳结果。DNA 片段中 *lac* 启动子的位置取决于所用限制酶。根据对称曲线，Wu 和 Crothers 推导出 *lac* 启动子内的 CAP-cAMP 结合位点就是弯曲中心。（*Source*：Wu，H. M.，and D. M. Crothers，The locus of sequence-directed and protein-induced DNA bending. *Nature* 308：511，1984.）

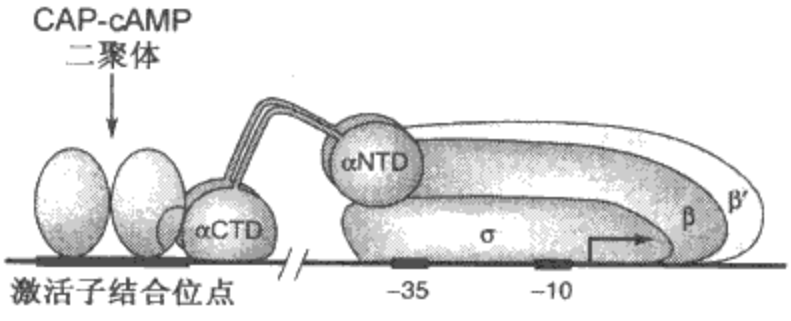


图 7.19　CAP-cAMP 激活 *lac* 操纵子转录的假设。CAP-cAMP 二聚体（紫色）结合靶 DNA 位点，αCTD（红色）与 CAP 蛋白上的特异位点（棕色）相互作用，增强了聚合酶与启动子间的结合。（*Source*：Adapted from Busby，S. and R. H. Ebright，Promoter structure，promoter recognition，and transcription activation in prokaryotes，*Cell* 79：742，1994.）

合启动子复合体，然后由闭合启动子复合体转换为开放启动子复合体。CAP-cAMP 通过与 RNA 聚合酶的 αCTD 发生蛋白质-蛋白质相互作用而招募 RNA 聚合酶与启动子的结合，CAP-cAMP 与靶 DNA 结合后引起 DNA 发生约 100°弯曲。

表 7.2　CAP-cAMP 对 *lac*P1 转录的激活

酶	转录本（cpm）				P1/UV5/%		激活（倍数）
	−cAMP-CAP		+cAMP-CAP				
	P1	UV5	P1	UV5	−cAMP-CAP	+cAMP-CAP	
α-WT	46	797	625	748	5.8	83.6	14.4
α-256	53	766	62	723	6.9	8.6	1.2
α-235	51	760	45	643	6.7	7.0	1.0

7.2　*ara* 操纵子

我们已经提及过，*E. coli* 的 *ara* 操纵子是细菌的另一类代谢物阻遏操纵子，能够编码代谢阿拉伯糖所需的酶。与 *lac* 操纵子相比，*ara* 操纵子具有一些较为有趣的特征。第一，存在 $araO_1$ 和 $araO_2$ 两个操纵基因，其中 $araO_1$ 调节控制基因 *araC* 的转录，而 $araO_2$ 位于其所控制的启动子 P_{BAD} 上游远处的 −265～−294，发挥控制转录的功能。第二，CAP 结合位点位于 *ara* 启动子上游约 200bp 处，CAP 具有促进转录的功能。第三，*ara* 操纵子具有由 AraC 蛋白介导的负调控系统。

ara 操纵子的阻遏环

$araO_2$ 如何调控位于其下游 250bp 的启动子转录？最为合理的解释是这两个位点（操纵基因和启动子）之间的 DNA 发生环化，如图 7.20a 所示。事实上，我们已经掌握了有关 DNA 发生环化的证据。Robert Lobell 和 Robert Schleif 发现，如果将具有整数双螺旋转角的 DNA 片段（10.5 的倍数）插入操纵基因与启动子之间，操纵基因仍然具有正常的功能。然而，如果插入非整数双螺旋转角（如 5bp 和 15bp）的 DNA 片段时，操纵基因就不能行使其功能了。这与下列一般观点是一致的，即只要 2 个蛋白质结合位点位于 DNA 双螺旋的同一侧，双链 DNA 分子就可以通过环化而使这 2 个蛋白质结合位点彼此靠近。然而，DNA 分子不可能扭转 180°而使不同侧面的两个位点拉到同一面侧，然后再通过环化使之发生相互作用（图 7.20）。从这一点上看，DNA 就像一段僵硬的晾衣线，可以弯曲，但不易扭曲。

图 7.20 所描述的简单模型中假定蛋白质首先与两个相距较远的 DNA 位点结合，然后通过蛋白质间的相互作用而使 DNA 环化。然而，Lobell 和 Schleif 研究发现，实际情况要比模型复杂得多。事实上，*ara* 操纵子的调控蛋白 AraC 是个双功能蛋白质，既可作为正调控蛋白又可作为负调控蛋白，其在操纵子上具有 3 个结合位点：位于远上游的 $araO_2$、定位在 −106 和 −144 之间的 $araO_1$ 以及由 2 个半位点 $araI_1$（−56～−78）和 $araI_2$（−35～−51）组成的 *araI*。每个位点可结合 1 个 AraC 单体。*ara* 操纵子包含 *araA*～*araD* 4 个结构基因，也称为 *araCBAD* 操纵子。其中 *araA*、*araB* 和 *araD* 编码阿拉伯糖代谢酶，从启动子 $araP_{BAD}$ 处向右转录。另一个基因 *araC* 编码调控蛋白 AraC，从启动子 $araP_C$ 处向左转录。

在阿拉伯糖缺乏时，生物体不需要 *araBAD* 基因的表达产物参与糖类代谢，因此 AraC 蛋白执行负调控因子作用，结合在 $araO_2$ 和 $araI_1$ 位点，引起两位点间的 DNA 环化，使操纵子受阻遏（图 7.21b）。当阿拉伯糖存在时，阿拉伯糖与 AraC 蛋白结合使其构象发生改变，不能与 $araO_2$ 结合，但能结合 $araI_1$ 和 $araI_2$，导致阻遏环（repression loop）解体，操纵子去阻遏（图 7.21c）。与 *lac* 操纵子一样，去阻遏并不是正调控的全部，还有 CAP-cAMP 复合体介导的正调控。图 7.21c 显示，该复合体结合在 $araP_{BAD}$ 启动子的上游。CAP-cAMP 在 $araP_{BAD}$ 启动子较远处结合却能控制转录，DNA 环化可能是合适的解释。DNA 成环使 CAP 与 RNA 聚合酶接触，进而促进 RNA 聚合酶与启动子的结合。

ara 操纵子阻遏环的证据

ara 操纵子阻遏环模型的证据是什么？首先，Lobell 和 Schleif 通过凝胶电泳研究发现，在阿拉伯糖缺乏时，AraC 能够引起 DNA 环化。在实验过程中，研究者并没有使用 *E. coli* 的整个 DNA 基因组，而是使用了称为微环（minicircle）的长 404bp 的超螺旋环形 DNA 分子。微环含有 $araO_2$ 和 $araI_1$ 位点，两者

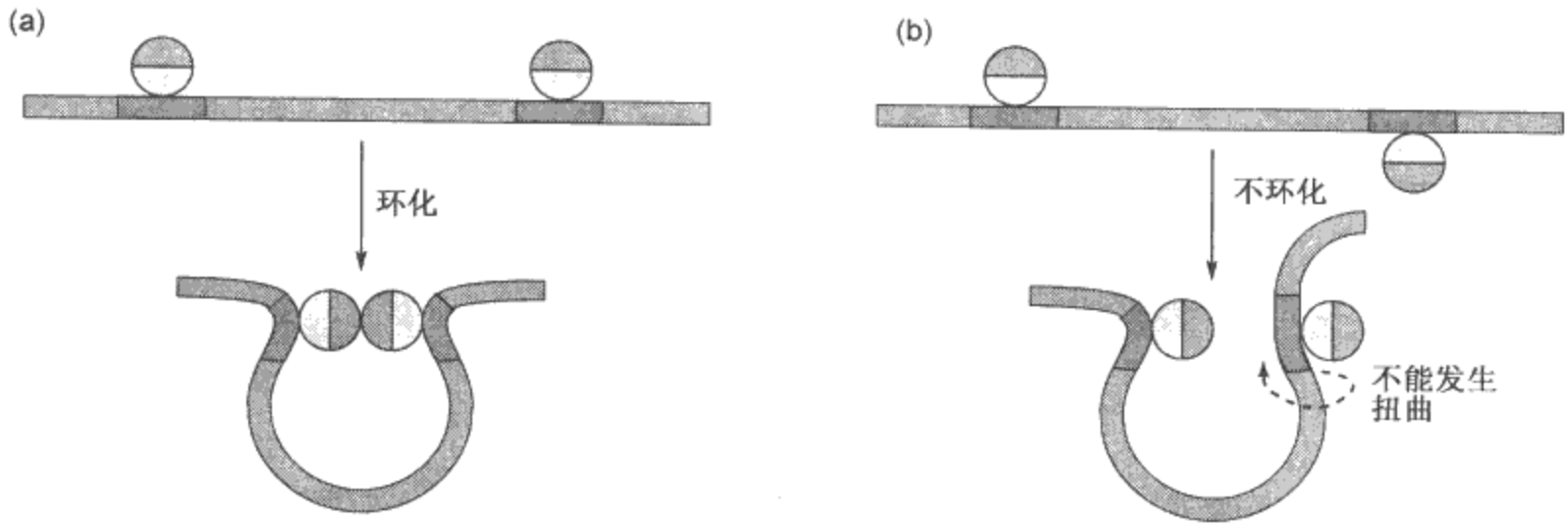


图 7.20　蛋白质必须结合在 DNA 的同一侧面才能通过 DNA 环化而发生相互作用。（a）具有 DNA 结合域（黄色）和蛋白质-蛋白质相互作用结构域（蓝色）的 2 个蛋白质与其靶 DNA 位点（红色）结合，处于 DNA 双螺旋的同一个侧面，两个结合位点间的 DNA 无须扭曲即可环化，从而使两个蛋白质发生相互作用；（b）两个蛋白质的结合位点位于 DNA 双螺旋的相反侧面，DNA 没有足够的柔韧性通过扭曲而使两个位点相互靠近，因此结合在 DNA 分子化的两个蛋白质不能发生相互作用。

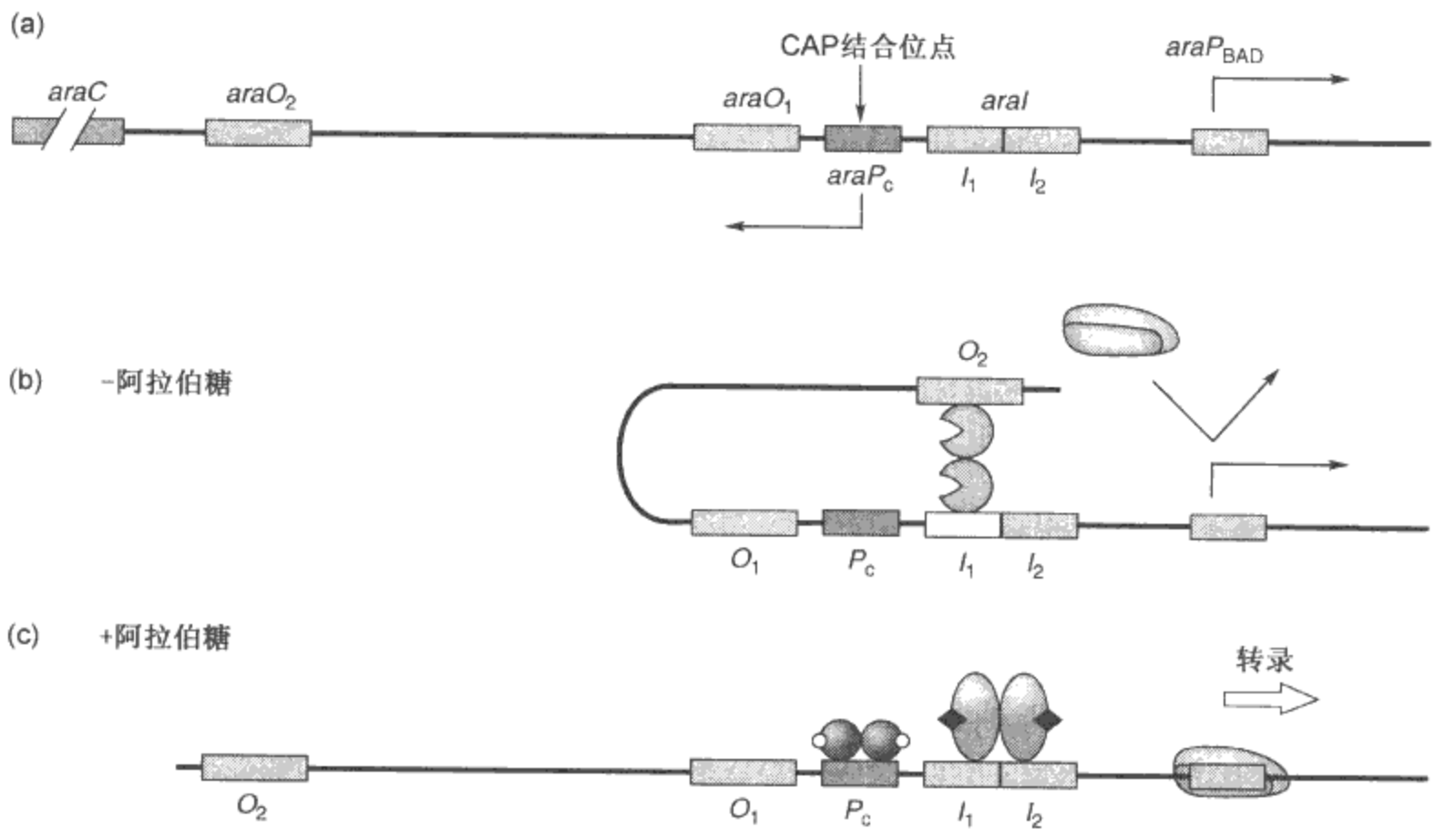


图 7.21　*ara* 操纵子的调控。（a）*ara* 操纵子调控区图谱。在启动子 $araP_{BAD}$ 上游区域有 4 个调控蛋白 AraC 的结合位点（$araO_2$、$araO_1$、*ara* I_1 和 *ara* I_2）。*ara* C 基因的转录在 $araP_C$ 起始由右向左进行。（b）负调控。当阿拉伯糖缺乏时，AraC 蛋白单体（绿色）结合在 $araO_2$ 和 ara_{I1} 位点上，引起 DNA 弯曲，阻止 RNA 聚合酶（红色和蓝色）靠近启动子。（c）正调控。阿拉伯糖（黑色）与 AraC 蛋白结合使其构象发生改变，以二聚体形式与 $araI_1$ 和 $araI_2$ 结合，但失去了与 $araO_2$ 结合的能力，由此打开启动子（粉色）使 RNA 聚合酶结合。如果葡萄糖缺乏，CAP-cAMP 复合体（紫色和黄色）具有较高浓度足以占据 CAP 结合位点，促使聚合酶与启动子紧密结合，激活操纵子。

相距 160bp。然后向体系中添加 AraC 蛋白，利用环化的超螺旋 DNA 其电泳迁移率比未环化的同一 DNA 分子电泳迁移率快的事实，分析加入 AraC 后微环的环化情况。图 7.22 是其中一个分析实验的结果，对比泳道 1 和 2 可以看出，加入 AraC 产生了新的高迁移速率条带，对应于环化的微环。

该实验结果还表明，成环的稳定性取决于 AraC 同时与 $araO_2$ 和 ara_I 位点的结合。Lobell 和 Schleif 分别用野生型微环 DNA、$araO_2$ 位点突变的微环 DNA 以及 ara_I 位点突变的微环 DNA 制备了成环复合体，

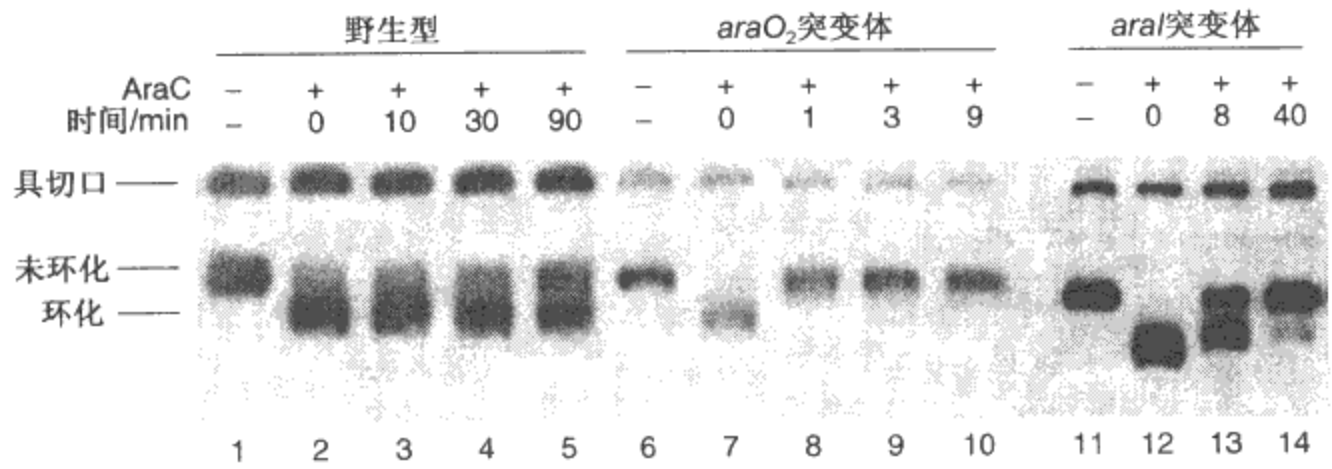


图 7.22　$araO_2$ 和 ara_I 突变对 *AraC* 成环复合体稳定性的影响。如图上方所示，Lobell 和 Schleif 制备了标记的微环 DNA 分子（小的环形 DNA 分子），分别含有野生型或突变型 AraC 蛋白结合位点。然后向体系中添加 AraC 蛋白，使之与标记 DNA 形成复合体。再加入过量的含有 ara_I 位点的未标记 DNA 作为竞争者，设定不同反应时间。最后将蛋白质-DNA 复合体进行凝胶电泳处理，检测 DNA 是否发生环化。成环 DNA 分子因比未成环 DNA 分子具有较多的超螺旋，因而电泳迁移速率快。有竞争者时，即使 90min 后，野生型 DNA 仍然保持成环复合体。相反，AraC 与突变体 DNA 解离，成环复合体很快就会消失。携带 $araO_2$ 突变 DNA 的成环复合体持续存在的时间少于 1min，而携带 *araI* 突变 DNA 的成环复合体在 10min 内有一半发生解离。（*Source*：Lobell，R. B. and Schleif，R. F.，DNA looping and unlooping by AraC protein. *Science* 250（1990），f. 2，p. 529. © AAAS.）

然后向体系中添加过量未标记的野生型微环 DNA，观察每种成环复合体的解体情况。泳道 3～5 显示在 90min 内，大约只有 50%的成环复合体转化为成环的野生型微观结构，说明野生型成环复合体的半衰期约为 100min。与之相反，$araO_2$ 突变微环从成环转换为开环的时间不到 1min（比较泳道 7 和 8）。*araI* 突变体的半衰期也很短，少于 10min。因此，$araO_2$ 位点和 ara_I 位点参与 AraC 引发的 DNA 环化，其中任一个发生突变都会影响 DNA 环化。

随后，Lobell 和 Schleif 研究表明，阿拉伯糖能够破坏阻遏环。在电泳前立即向体系中添加阿拉伯糖，结果在凝胶上对应成环 DNA 分子的条带消失，图 7.23 显示了该现象。如果去除阿拉伯糖，即使环被破坏了也能重新形成。用阿拉伯糖阻止 DNA 环化，然后用含过量竞争 DNA 及有或无阿拉伯糖的缓冲液稀释 DNA。含有阿拉伯糖的缓冲液能够维持 DNA 开环状态，而无阿拉伯糖的缓冲液中 DNA 稀释到一定程度后，开环的 DNA 可以重新环化。通常，利用电泳可检测 DNA 成环情况，图 7.24 为实验结果。

通过该实验可得出一个重要推论：1 个 AraC 二聚体足以保证 DNA 环的形成。因为最初的 AraC-DNA 复合体是在过量未标记 DNA 和阿拉伯糖存在时形成的，该条件确保了 1 个 AraC 二聚体结合在 ara_I 位点而不是 $araO_2$ 上。在过量未标记 DNA 阻止游离 AraC 二聚体与 $araO_2$ 位点结合的情况下也能成环，说明最初结合在 ara_I 上的 AraC 二聚体必须结合到 $araO_2$ 上才能引起 DNA 环化。Lobell 和 Schleif 用^{35}S 标记 AraC 蛋白、^{32}P 标记 DNA，然后测定成环复合体中每种同位素的含量。化学计量分析表明，每个 DNA 微环分子上结合了一个 AraC 二聚体。

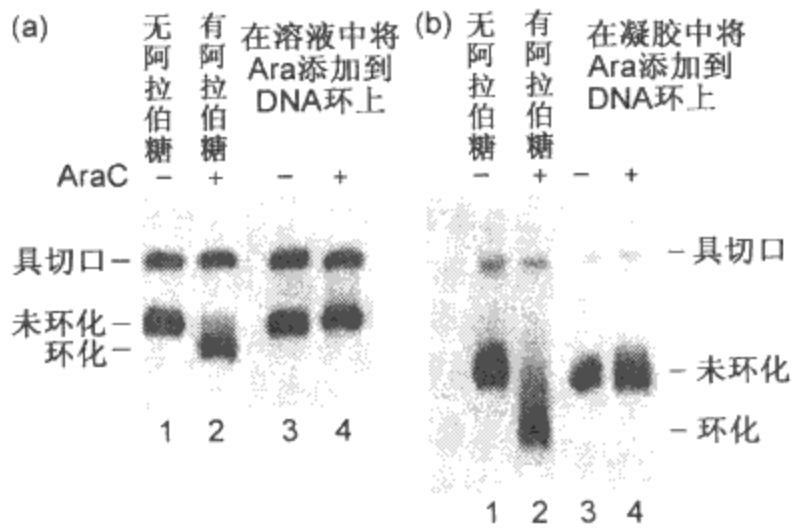


图 7.23　阿拉伯糖对 $araO_2$ 与 ara_I 位点间的 DNA 环的破坏。（a）Lobell 和 Schleif 在电泳前将阿拉伯糖加入已形成 DNA 环的体系中。无阿拉伯糖时，AraC 促进 DNA 成环（泳道 2），有阿拉伯糖时，由 AraC 促进形成的 DNA 环被破坏（泳道 4）。（b）电泳开始后将阿拉伯糖加入到凝胶中。同样，无阿拉伯糖时，AraC 能够促进 DNA 环的形成（泳道 2），有阿拉伯糖时，形成的 DNA 环被破坏（泳道 4）。Ara 表示阿拉伯糖。（*Source*：Lobell R. B.，and Schleif R. F.，DNA looping and unlooping by AraC protein. *Science* 250（1990），f. 4，p. 530. © AAAS.）

当 DNA 处于开环状态时，结合在 $araO_2$ 上的单体 AraC 会怎样呢？很显然，AraC 会与 $araI_2$ 位点结合。为证明这一点，Lobell 和 Schleif 首先通过**甲基化干扰实验**（methylation interference）证实，

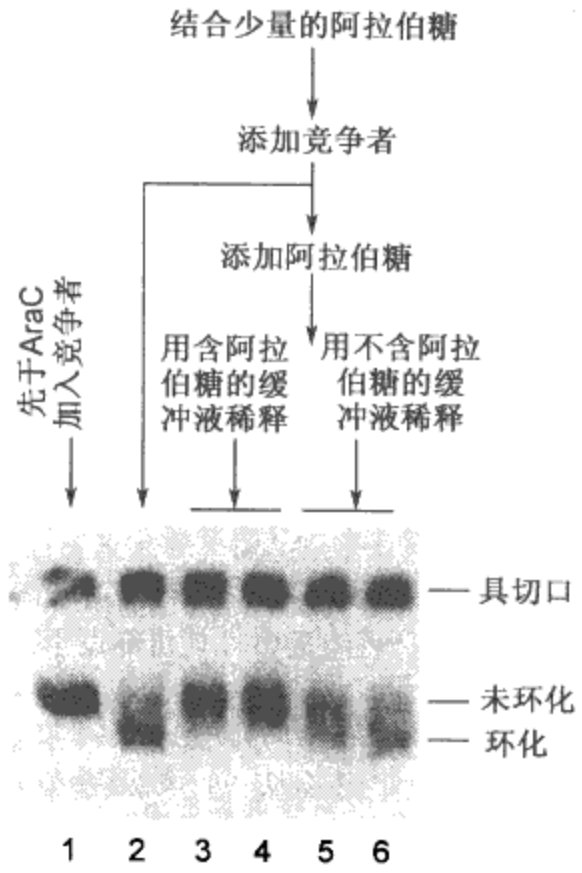


图 7.24　被阿拉伯糖打开的环的可逆性。Lobell 和 Schleif 用 AraC 和标记 DNA 形成环，然后加入过量未标记竞争 DNA，再加入阿拉伯糖，用不含或含阿拉伯糖的缓冲液稀释。实验操作流程见图上半部分。泳道 1，先于 AraC 加入竞争者，未检测到成环复合体。泳道 2，加入 AraC 后再加入竞争者，成环复合体是稳定的。泳道 3 和 4，加入阿拉伯糖和竞争剂，复合体不稳定，环断裂。泳道 5 和 6，用不含阿拉伯糖的缓冲液稀释体系中的阿拉伯糖，成环复合体重新形成。（*Source*：Lobell R. B.，and Schleif R. F.，DNA looping and unlooping by AraC protein. *Science* 250（1990），f. 5，p. 530. © AAAS.）

在开环时，AraC 与 $araI_1$ 结合，但不会与 $araI_2$ 结合。实验的策略是先使微环 DNA 发生部分甲基化修饰，然后加入 AraC 使 DNA 环化，电泳分离成环 DNA 和未成环 DNA，然后用特定试剂在甲基化位点处切割成环 DNA 和未成环 DNA。因为在这些重要位点的甲基化修饰会阻断 DNA 成环，所以在成环 DNA 中参与成环的重要位点不被甲基化，而在未成环 DNA 中这些位点被甲基化。事实上，在 $araI_1$ 位点内存在有 2 个碱基，其在未成环 DNA 分子中是高度甲基化的，而在成环 DNA 分子中只发生轻微甲基化修饰。与之相反，在 $araI_2$ 位点没有发现具有这种行为的碱基。因此，在成环状态下 AraC 不会与 $araI_2$ 结合。

Lobell 和 Schleif 通过突变实验进一步证实了上述结论。在成环状态，$araI_2$ 突变不影响 AraC 与 DNA 的结合，但在开环状态则有强烈影响。图 7.25 显示了 $araI_2$ 位点的突变效应。通过 AraC 的结合，使野生型或 $araI_2$ 突变型微环 DNA 发生环化，然后用阿拉伯糖使形成的 DNA 环打开。随后用 *Hind* III 处理使微环 DNA 线性化，电泳检测 AraC 是否仍与 DNA 结合。如果 AraC 仍结合在 DNA 上，线性 DNA 分子的电泳迁移率会变慢，移动滞后。因此，这也是凝胶迁移率变动分析的一种方式（第 5 章）。正如图 7.25 所示，当 DNA 环被破坏后，AraC 与野生型 DNA 结合，但在相似条件下不与 $araI_2$ 突变型 DNA 结合。因此我们推断，在开环状态下，$araI_2$ 是 AraC 与 DNA 结合所必需的，因此在这些条件下，$araI_2$ 位点能够与 AraC 结合。

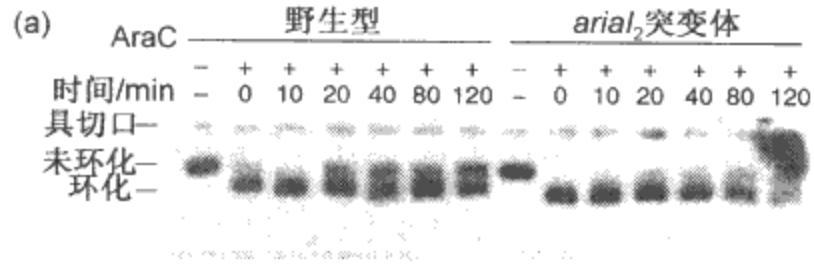


图 7.25　在开环复合体而非成环复合体中，AraC 能够与 $araI_2$ 结合。（a）按图 7.22 的方法，Lobell 和 Schleif 分析了野生型和 $araI_2$ 突变型 DNA 成环复合体的解离速率。结果显示两种复合体之间没有差异，表明成环状态下，AraC 与 $araI_2$ 没有结合。（b）检测在开环状态下 AraC 和 $araI_2$ 的结合。将 AraC 结合到标记的野生型或 $araI_2$ 突变型 DNA 上，用竞争 DNA 和阿拉伯糖处理 3～5min，打开环。如图顶部所示，用 *Hind* III 线性化 DNA，然后电泳产物，游离的线性 DNA 迁移率相对较快，易与结合 AraC 的线性 DNA 区分开。在开环状态，野生型 DNA 仍与 AraC 结合，但是 $araI_2$ 突变型 DNA 不结合 AraC。因此，$araI_2$ 与 AraC 在开环状态下结合，在成环时不结合。（*Source*：Lobell R. B.，and Schleif R. F.，DNA looping and unlooping by AraC protein. *Science* 250（1990），f. 7，p. 531. © AAAS.）

基于这些数据，研究者提出了如图 7.21b 和 c 所示的 AraC-DNA 相互作用模型。1 个 AraC 二聚体同时与 $araI_1$ 和 $araO_2$ 相互作用而导致 DNA 环化。阿拉伯糖与 AraC 结合后引起 AraC 构象的改变，使其失去对 $araO_2$ 的亲和力，转而与 $araI_2$ 结合，最终导致 DNA 环解体。

araC 的自主调节

到目前为止，我们只提及了 $araO_1$ 的作用，$araO_1$ 并没有参与对 *araBAD* 转录的阻遏，只是使 AraC 调节自身的合成。图 7.26 显示了 *araC*、P_C 和 $araO_1$ 的相对位置。*araC* 的转录从 P_C 处起始从右向左进行，$araO_1$ 处于调控 *araC* 转录的位置上。随着 AraC 蛋白含量的不断增加，AraC 结合在 $araO_1$ 上，抑制 *araC* 的转录，防止阻遏物过多积累。这种蛋白质控制自身合成的机制称为自主调节（autoregulation）。

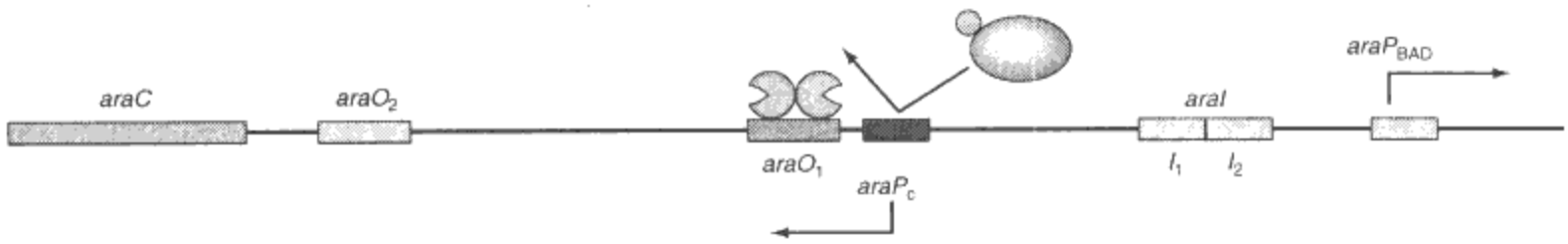


图 7.26 *araC* 的自主调节。AraC 蛋白（绿色）结合到 $araO_1$ 位点上，阻止 P_C 对 *araC* 的转录，无论阿拉伯糖与 AraC 结合与否，或者说不论 *ara* 操纵子的控制区发生环化与否，这种阻遏事件均可能发生。

小结 *ara* 操纵子受 AraC 蛋白调控。AraC 能够使相距 210bp 的 $araO_2$ 和 $araI_1$ 两位点之间的 DNA 发生环化，从而阻遏 *ara* 操纵子的表达。阿拉伯糖的结合能够松弛 AraC 蛋白与 $araO_2$ 之间的结合，进而与 $araI_2$ 结合，使 *ara* 操纵子去阻遏。CAP 和 cAMP 通过与位于 *araI* 上游的位点结合而促进 *ara* 操纵子转录。AraC 通过与 $araO_1$ 位点结合而阻止 *araC* 基因由右向左的转录进而调控自身的合成。

7.3 *trp* 操纵子

E. coli 的 ***trp***（发音为“trip”）操纵子能够编码合成色氨酸所需要的酶。与 *lac* 操纵子一样，*trp* 操纵子受阻遏物的负调控，而两者具有根本性的区别。*lac* 操纵子编码**分解代谢**（catabolite）底物的酶，只有当被分解的底物（如乳糖）存在时，操纵子才被打开。而 *trp* 操纵子则编码参与底物**合成代谢**（anabolic）的酶，当底物存在时，操纵子的表达是关闭的。当色氨酸浓度升高时，生物体不再需要 *trp* 操纵子产物，*trp* 操纵子应该被阻遏。这正是实际发生的反应。此外，*trp* 操纵子还具有 *lac* 操纵子所没有的衰减调控（attenuation）机制。

色氨酸在 *trp* 操纵子负调控中的作用

图 7.27 显示了 *trp* 操纵子的大致结构。5 个基因编码的酶将色氨酸前体分支酸（chorismic acid）转化为色氨酸。在 *lac* 操纵子中启动子和操纵基因位于结构基因的上游，*trp* 操纵子的结构次序也是如此，但是 *trp* 操纵基因完全位于 *trp* 启动子内部，在 *lac* 操纵子中启动子和操纵基因只是相邻而已。

在 *lac* 操纵子的负调控机制中，细胞通过微量异乳糖的出现而感应乳糖的存在。实质上，异乳糖作为诱导物能够使阻遏物从 *lac* 操纵基因上解离下来，使 *lac* 操纵子去阻遏。在 *trp* 操纵子的负调控机制中的，大量色氨酸供应意味着细胞无须再消耗更多能量去合成这种氨基酸，即高浓度色氨酸是关闭 *trp* 操纵子的信号。

细胞如何感知色氨酸的存在？实质上，色氨酸能够帮助 *trp* 阻遏物与操纵基因结合。在缺乏色氨酸时，只有无活性的无辅基阻遏物蛋白（aporepressor）的存在，而无 *trp* 阻遏物的存在。当无辅基阻遏物蛋白与色氨酸结合后，其构象发生改变，增强了对 *trp* 操纵基因的亲和性（图 7.27b）。就像我们在讨论 *lac* 操纵子调控机制所遇到的情况一样，这是另一个同分异构转换的事例。无辅基阻遏物蛋白与色氨酸结合后形成**色氨酸阻遏物**（*trp* repressor），因此，色氨酸又称为**辅阻遏物**（corepressor）。当细胞中色氨酸浓度较高时，大量的辅阻遏物与无辅基阻遏物蛋白结合，形成有活性的 *trp* 阻遏物，使操纵子被阻遏。当细胞中色氨酸水平降低时，色氨酸与无辅基阻遏物蛋白解离，使无辅基阻遏物蛋白转变为无活性的构象，导致阻遏物-操纵基因复合体解离，操纵子解阻遏。在第 9 章，我们将分析无辅基阻遏物蛋白与色氨酸结合时构象变化的实质，并了解构象的转换对于阻遏物与操纵基团结合的重要性。

小结 在某种意义上，*trp* 操纵子的负调控具有与 *lac* 操纵子的负调控互为映像的特征。*lac* 操纵子对诱导物的应答会引起阻遏物与操纵基因的解离，从而使操纵子去阻遏。而 *trp* 操纵子则在细胞内色氨酸浓度很高时，能够应答包括辅阻遏物色氨酸在内的阻遏物。辅阻遏物与无辅基阻遏蛋白结合，使之构象发生改变，从而更好地与 *trp* 操纵基因结合，导致操纵子被阻遏。

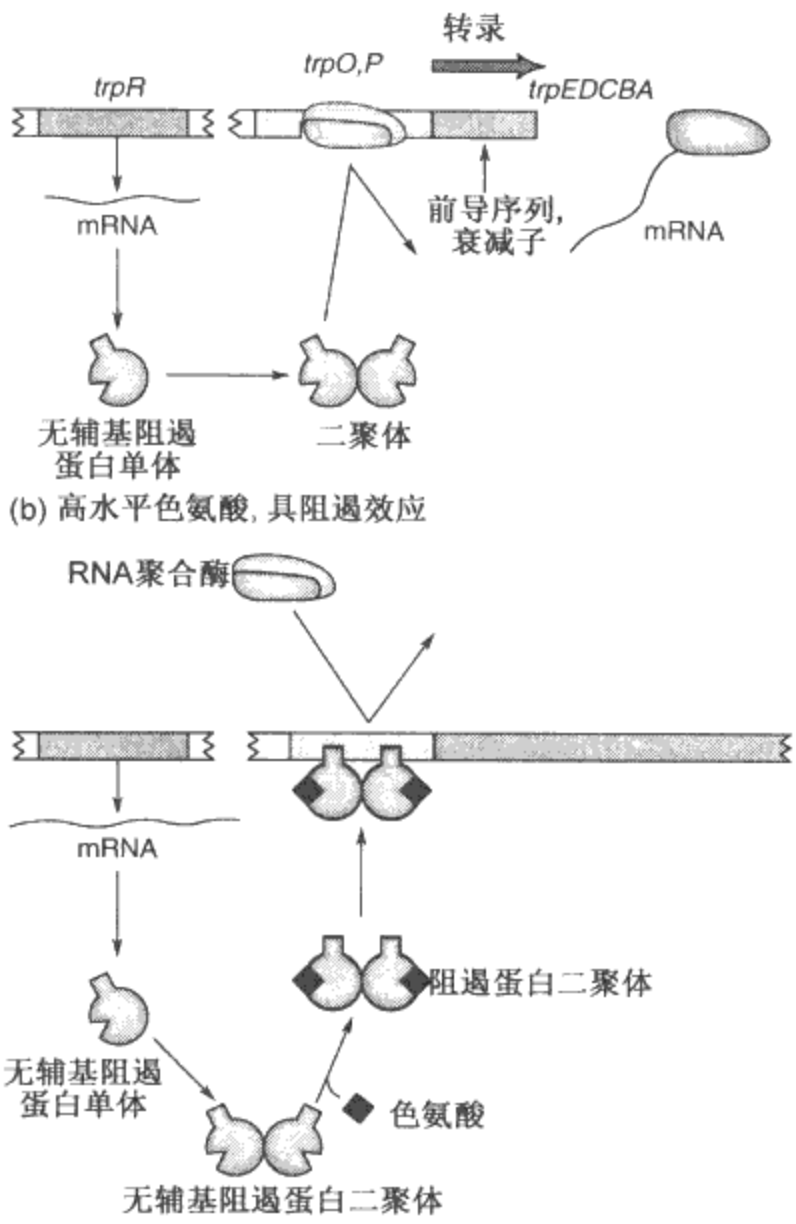


图 7.27　*trp* 操纵子的负调控。（a）去阻遏。RNA 聚合酶（红色和蓝色）与 *trp* 启动子结合，起始结构基因（*trpE*、*trpD*、*trpC*、*trpB*、*trpA*）的转录。无色氨酸时，阻遏物蛋白（绿色）不能与操纵基因结合。（b）阻遏。色氨酸作为辅阻遏物（黑色）与无活性的无辅基阻遏物蛋白结合，使之成为有活性的阻遏物，具有与 *trp* 操纵基因结合的构型，从而阻止 RNA 聚合酶与启动子结合，因此无转录发生。

衰减作用对 *trp* 操纵子的调控

除了前面所讲的标准负调控模式外，*trp* 操纵子还具有另一种称为**衰减作用**（attenuation）的调控机制。为什么 *trp* 操纵子需要这种额外的调控机制？其可能的原因在于 *trp* 操纵子的阻遏作用实质上是一种弱调控机制，其调控作用远弱于 *lac* 操纵子的阻遏作用。因此，即使有阻遏物存在的情况下，*trp* 操纵子仍能发生大量的转录事件。事实上，在只有阻遏调控的衰减子突变体中，*trp* 操纵子处于完全阻遏时的转录活性仅是完全阻遏时转录活性的 1/70。而衰减作用系统可以在 10 倍范围内对 *trp* 操纵子的转录活性进行调控。因此在阻遏调控系统和衰减调控系统共同作用下，可以在 700 倍左右的范围内对 *trp* 操纵子的转录活性进行调控，即从完全失活到完全激活其转录效率相差 700 倍，70 倍（阻遏）×10 倍（衰减）＝700 倍。这是十分经济有效的，因为色氨酸的合成需要消耗大量的能量。

下面我们来具体了解一下衰减作用的机制。图 7.27 显示，在操纵基因和第一个结构基因 *trpE* 之间具有 ***trp* 前导区**（*trp* leader）和 ***trp* 衰减子**（*trp* attenuator）2 个位点。图 7.28 详尽介绍了前导区-衰减子的特征。其作用是在色氨酸含量相对丰余时，衰减或弱化 *trp* 操纵子的转录，衰减子通过促使转录提前终止而发挥作用。换句话说，即使在色氨酸浓度较高时，*trp* 操纵子也能起始转录，但有 90% 的转录在衰减子区发生终止。

提前终止的原因是衰减子含有转录终止信号（终止子）：反向重复序列及其随后连续排列的 8 个 A-T 碱基对。由于存在反向重复序列，该区域的转录物可以通过分子内碱基配对，而形成如图 7.29 所示发夹结构。与第 6 章学习内容的一样，转录物发夹结构之后的一串 U 能够降低转录物与 DNA 结合的稳定性，从而导致转录的终止。

小结　除了阻遏蛋白-操纵基因调控系统外，*trp* 操纵子还利用额外的衰减调机制调控基因表达。在合成产物充足时，衰减作用通过使操纵子转录提前终止而发挥作用。

衰减作用的失效

当色氨酸缺乏时，*trp* 操纵子必须能够被激活，这意味着细胞必须以某种方式使衰减调控机制失效。Charles Yanofsky 提出假设，认为有某种物质阻止发夹形成，破坏了终止信号，从而使衰减作用终止，*trp* 操纵子的转录得以继续进行。观察图 7.30a 可见，在前导区转录物末端附近可形成两个潜在的发夹结构。其中第二个发夹结构中含有终止信号，与其后一连串的 U 紧邻。另外，在该区域内形成发夹结构的排列组合方式并不是唯一的，图 7.30c 显示了形成发夹结构的另一种组合方式，该发夹由第一种组合方式中形成的 2 个发夹各贡献一个元件组成。图 7.30b 概括了发夹结构形成的方式，最初的 2 个发夹结构的基部元件分别标记为 1、2、3、4。第一个发夹包含元件 1 和 2，第二个发夹包含元件 3 和 4，而由第二种方式形成的发夹结构包含元件 2 和 3。这就意味着元件 2 和 3 配对形成第二种发夹结构（图 7.30b）后，就能阻止元件 1 和 2 配对、3 和 4

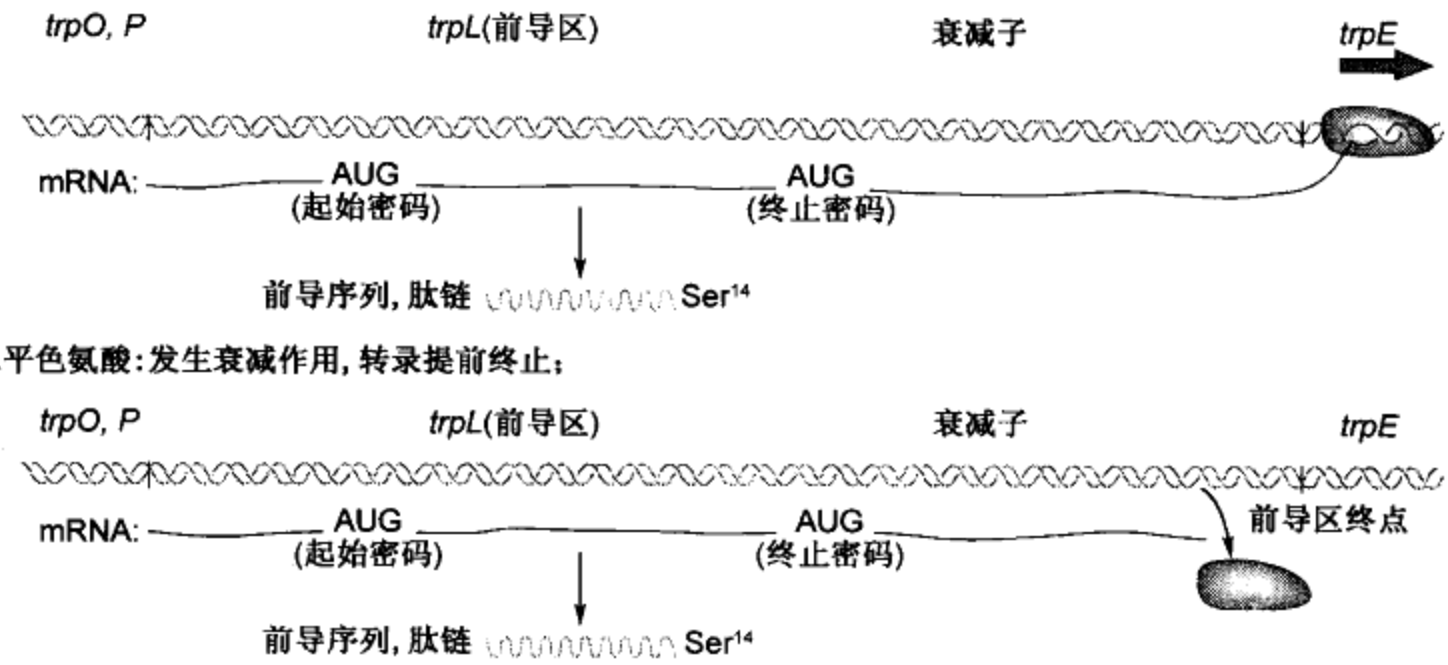


图 7.28　*trp* 操纵子的衰减作用。（a）在色氨酸低浓度条件下，RNA 聚合酶（红色）通读衰减子，结构基因被转录；（b）在高色氨酸浓度条件下，衰减子会引起转录提前终止，结构基因不能被转录。

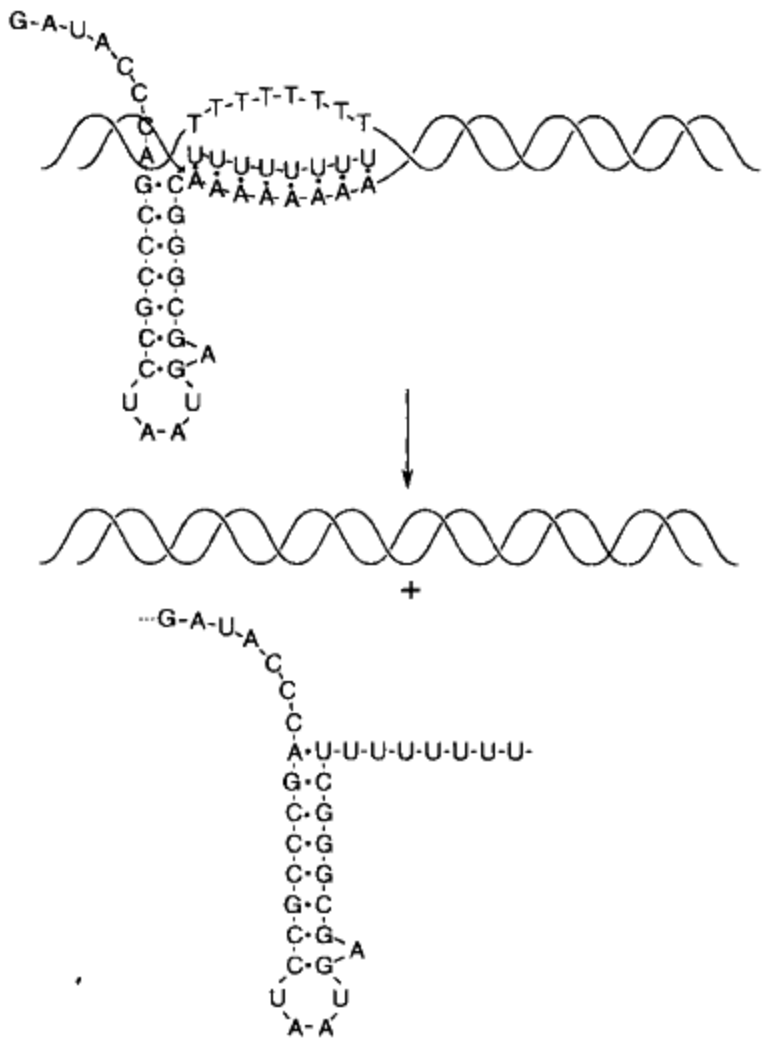


图 7.29　衰减作用机制。前导区-衰减子区（蓝色）的转录物含有一串 U 和反向重复序列，能够形成发夹。当 RNA 聚合酶在 U 串处停顿时，发夹形成，释放转录物，所以转录到达 *trp* 基因之前发生了终止。

配对，这样就不可能通过第一种方式形成 2 个发夹结构，包括与一连串 U 相邻的那个发夹结构，而一连串的 U 是终止子的必需部分（图 7.30a）。

双发夹所包含的碱基对比单发夹要多，所以双发夹结构更稳定。那么，为什么要形成稳定性差的单发夹呢？从前导区的碱基序列特征中可以找到相应的线索，如图 7.31 所示。该序列的显著特征是在元件 1 中存在 2 个串联在一起的色氨酸密码子（UGG）。这似乎是正常现象，但在大多数蛋白质中，色氨酸属于稀有氨基酸，平均每 100 个氨基酸中才会出现一个，所以两个色氨酸密码子并列串联在一起的概率很小，在 *trp* 操纵子中出现两个并列氨基酸密码子并列串联在一起的事实很值得深思。

在细菌中，转录和翻译是同时发生的。只要 *trp* 前导区起始转录，核糖体就开始翻译已转录的 mRNA。试想处于色氨酸缺乏情况下，核糖体在试图翻译前导区内的 2 个色氨酸密码子时将做如何处理（图 7.32a）？当色氨酸供应不足时，但对两个位点的翻译又需要色氨酸。十有八九核糖体无法立即满足翻译的需求，因此核糖体会在一个 Trp 密码子处停顿下来。那么停顿下来的核糖体占据在前导区转录物序列的什么部位呢？正好在形成第一个发夹的元件 1 上，庞大的核糖体依附在 RNA 位点上有效地阻止了元件 1 与元件 2 的配对，游离的元件 2 便与元件 3 配对形成单发夹结构。因为第二个发夹（元件 3 和 4）不能形成，转录不能终止，衰减作用失效。这正是生物体所需要的，因为色氨酸缺乏时 *trp* 操纵子应该被转录。

值得注意的是，该机制涉及转录和翻译的偶联，且翻译影响转录。真核生物没有这样的机制，因为真核生物的转录和翻译发生在细胞的不同区室内。细菌的衰减机制也有赖于转录和翻译以大致相同速度进行，如果 RNA 酶的转录速率快于核糖体的翻译速率，那么，在核糖体在到达色氨酸密码子处并发生停留之

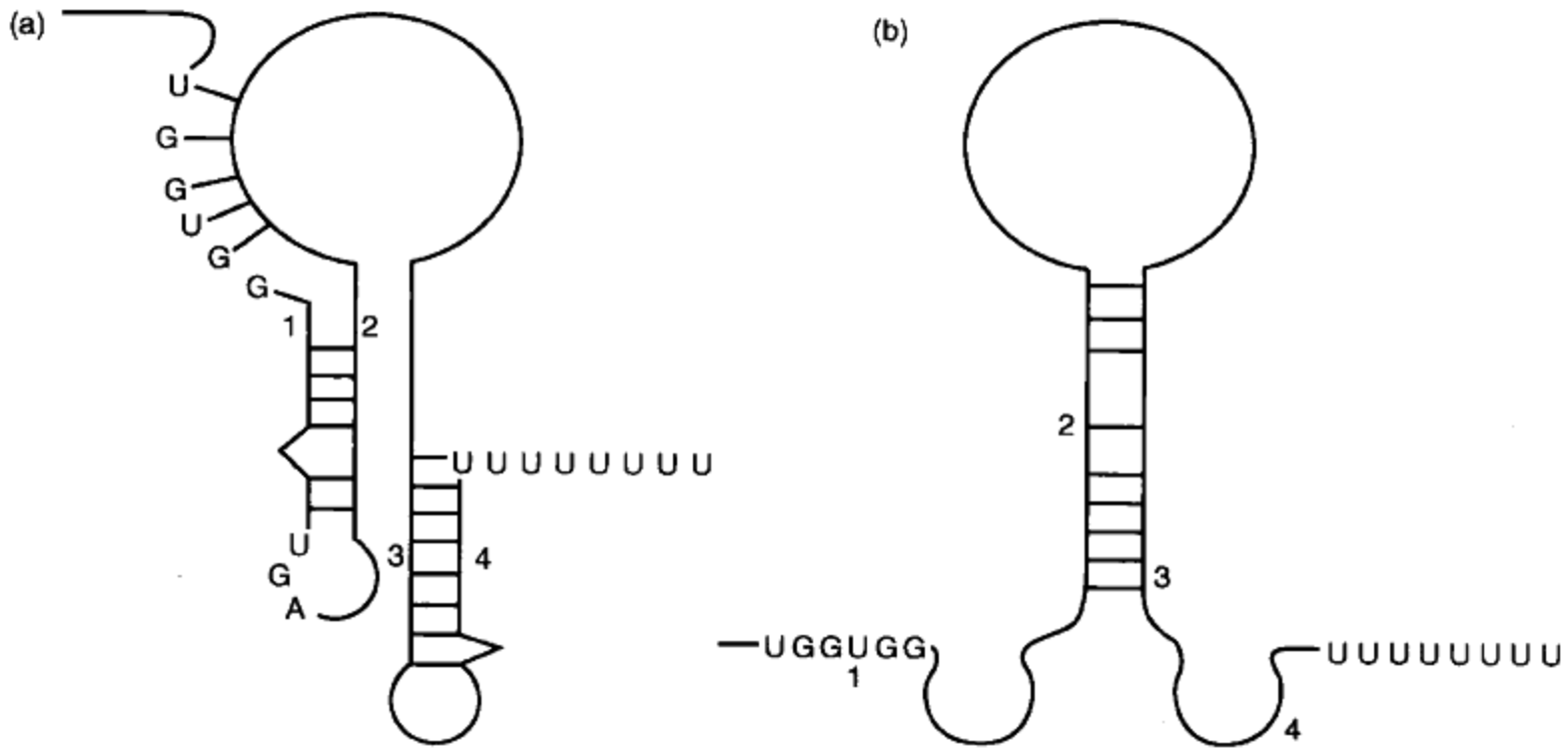


图 7.30　前导区-衰减子转录物的两种结构。（a）具有 2 个发夹环的稳定结构。（b）只有 1 个发夹环的亚稳定结构。图底部所示 RNA 的弯曲形状并不代表分子的实际形状，之所以这样绘画是为了节约篇幅。参与碱基配对的序列（1～4）用彩色线条标注，图（a）和图（b）的颜色标注相同。

Met Lys Ala Ile Phe Val Leu Lys Gly Trp Trp Arg Thr Ser 终止密码
pppA---AUGAAAGCAAUUUUCGUACUGAAAGGUUGGUGGCGCACUUCCUGA

图 7.31　前导区序列。显示前导区转录物的部分序列及其编码的前导肽序列。注意，2 个 Trp 密码子串联在一起（蓝色）。

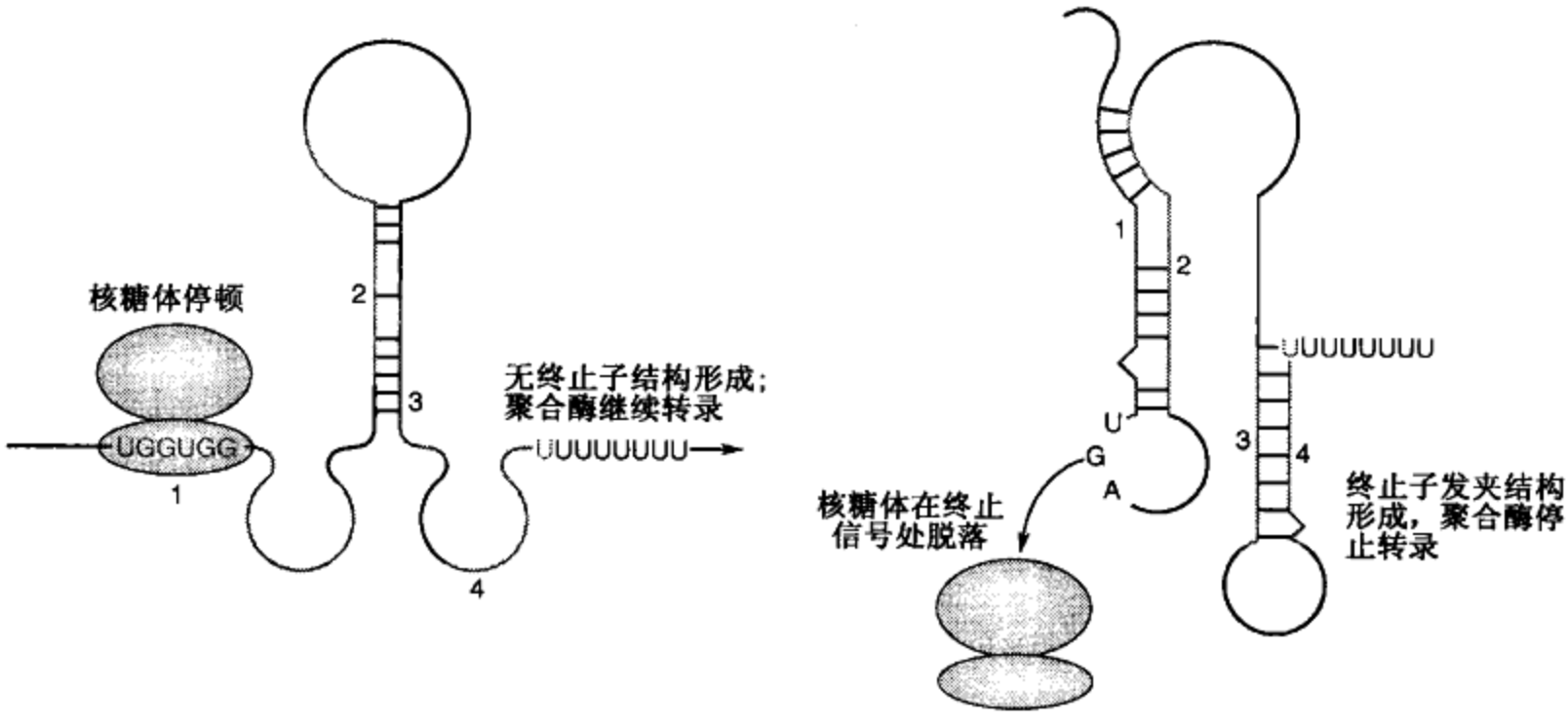


图 7.32　衰减作用的克服。（a）在色氨酸饥饿条件下，核糖体（黄色）停顿在色氨酸密码子处，阻止元件 1（红色）和 2（蓝色）间的碱基配对，这样就只能形成一个无终止子的单发夹结构，因此不会产生衰减作用。（b）色氨酸充足时，核糖体通读 2 个色氨酸密码子，翻译形成前导肽，并在翻译终止信号（UGA）处脱离 mRNA。因此，核糖体不干扰前导区转录物的分子内碱基配对，形成更稳定的双发夹结构，其中包含终止子，所以能够发生衰减作用。

前，聚合酶可能已完成了对衰减子的转录。

如果核糖体停留在前导区的起始端，*trp* 操纵子合成的多顺反子 mRNA 如何能够被翻译成蛋白质？由于 mRNA 上对应的每个基因都有各自的翻

译起始信号（AUG），核糖体能够分别独立地识别这些信号，因此，前导区的翻译不影响 *trp* 基因的翻译。

另一方面，在色氨酸充足的条件下，核糖体翻译前导区转录物的情况如下（图 7.32b）；双 Trp 密码子不成为核糖体对前导区转录物翻译的屏障，核糖体能够阅读元件 1 继续进行翻译，直到元件 1 和 2 之间的终止信号（UGA），然后从转录本上脱离下来。由于没有核糖体的干扰，前导区转录物能够形成 2 个发夹，完成终止子结构的形成，产生终止信号，使 RNA 聚合酶在到达 *trp* 基因之前停止了转录。由此可见，衰减系统能够感应色氨酸存在的丰余，避免细胞合成更多色氨酸合成酶。

除了 *trp* 操纵子外，*E. coli* 其他的操纵子也具有衰减机制。*E. coli his* 操纵子的前导区含有 7 个连续的组氨酸密码子，以此阻止核糖体对前导区的翻译。

小结 *E. coli trp* 操纵子的衰减机制能够在色氨酸充足的情况下发挥作用。当色氨酸供应受限时，核糖体在 *trp* 操纵子前导区的两个串联 trp 密码子处停顿。因为核糖体停顿时，*trp* 前导区正处于转录过程中，因此核糖体的停顿会影响 RNA 的折叠方式，特别是能够影响携带部分转录终止信号的发夹结构的形成，该转录终止信号能够引发衰减作用。因此，当色氨酸缺乏时，衰减作用失效，操纵子保持活化状态，这意味着，同阻遏作用一样，衰减系统通过感应色氨酸浓度水平而发挥调控作用。

7.4 核糖开关

在上一节介绍的基因表达调控事例中，细胞通过控制 mRNA 的 5′-非翻译区（UTR）的结构而调控基因表达（*E. coli trp* 操纵子 mRNA）。在该事例中，结合在 *trp* 5′-UTR 上的大分子集合体（核糖体）能够感应小分子（色氨酸）的浓度，从而通过改变 5′-UTR 的结构而调控转录。这是小分子（ligand 配体）通过大分子介导而影响基因表达的一个事例。

有关小分子通过直接作用于 mRNA 的 5′-UTR 而调控基因表达的事例也越来越多。在 mRNA 的 5′-UTR 中有些区域能够感应配体的结合而改变自身结构，进而调控基因的表达，mRNA 5′-UTR 中的这些元件就是**核糖开关**（riboswitch）。

核糖开关中能够结合配体的区域称为**适体**（aptamer）。适体是科学家在研究进化时首次发现的，利用体外快速复制的 RNA 来筛选与配体特异性紧密结合的短 RNA 序列。由于 RNA 在复制过程中会发生错误，产生了新的 RNA 序列，可从中筛选出与特定配体结合最理想的 RNA 序列。在体外实验过程中发现了许多适体，奇怪的是生物体为什么不利用这些适体？我们现在知道生物体是能够利用这些适体的。

核糖开关的经典范例是枯草杆菌的 *ribD* 操纵子，该操纵子控制维生素核黄素（riboflavin）及其产物之一黄素单核苷酸（flavin mononucleotide，FMN）的合成与转运。*rib* 操纵子的 5′-UTR 含有保守的 **RFN 元件**，该元件突变会破坏 FMN 对操纵子的正常调控。于是有人提出假设，认为 RFN 元件能够与应答 FMN 的蛋白质或 FMN 自身发生相互作用。

为验证 RFN 元件就是能够直接与 FMN 结合的适体的假设，Ronald Breaker 及同事采用了线内探测（in-line probing）技术来分析核糖开关。其原理是 RNA 磷酸二酯键的有效水解（断裂）需要在亲核试剂（水）的攻击下发生 180°重排，水解后形成游离的磷原子和末端带有羟基基团的 RNA 片段。无结构的 RNA 易于采取这种线内构型，但是二级结构（分子内碱基配对）的 RNA 则不能采取这种线内构型。所以，线性无结构 RNA 容易发生裂解，而有结构的 RNA 因具有较多的碱基对而不易发生裂解。

因此，Breaker 及同事在 FMN 有或无的条件下温育含 RFN 元件的标记 RNA 片段。图 7.33a 显示 RNA 片段在 FMN 有或无的条件下自发水解的情况是不同的，提示 FMN 可直接与 RNA 结合，引起 RNA 构型转换，这正是我们预期适体与配体结合的情形。

特别是 Breaker 及同事发现 FMN 的结合会保护某些特定的磷酸二酯键不易受裂解，而其他磷酸二酯键仍然保持其原有易裂解性（图 7.33b）。而且，在 FMN 浓度仅为 5nmol/L 时，磷酸二酯键易裂解性至少发生了 50%的改变，表明 RNA 片段与配体之间具有很高的亲和性。

在 FMN 存在时 RNA 易裂解性降低的特性表明 RFN 元件具有两种可变构型，如图 7.33c 所示。无 FMN 时，RFN 元件形成抗终止子结构，其发夹结构远离其后的 6 个 U。但 FMN 的存在使其构型转换，形成终止子结构，阻断操纵子的表达。这是合理的，因为 FMN 充足时，无须 *ribD* 操纵子表达，所以 FMN 引发的衰减作用有利于细胞节约能量。

为验证该假说，Breaker 及同事以包含 RFN 元

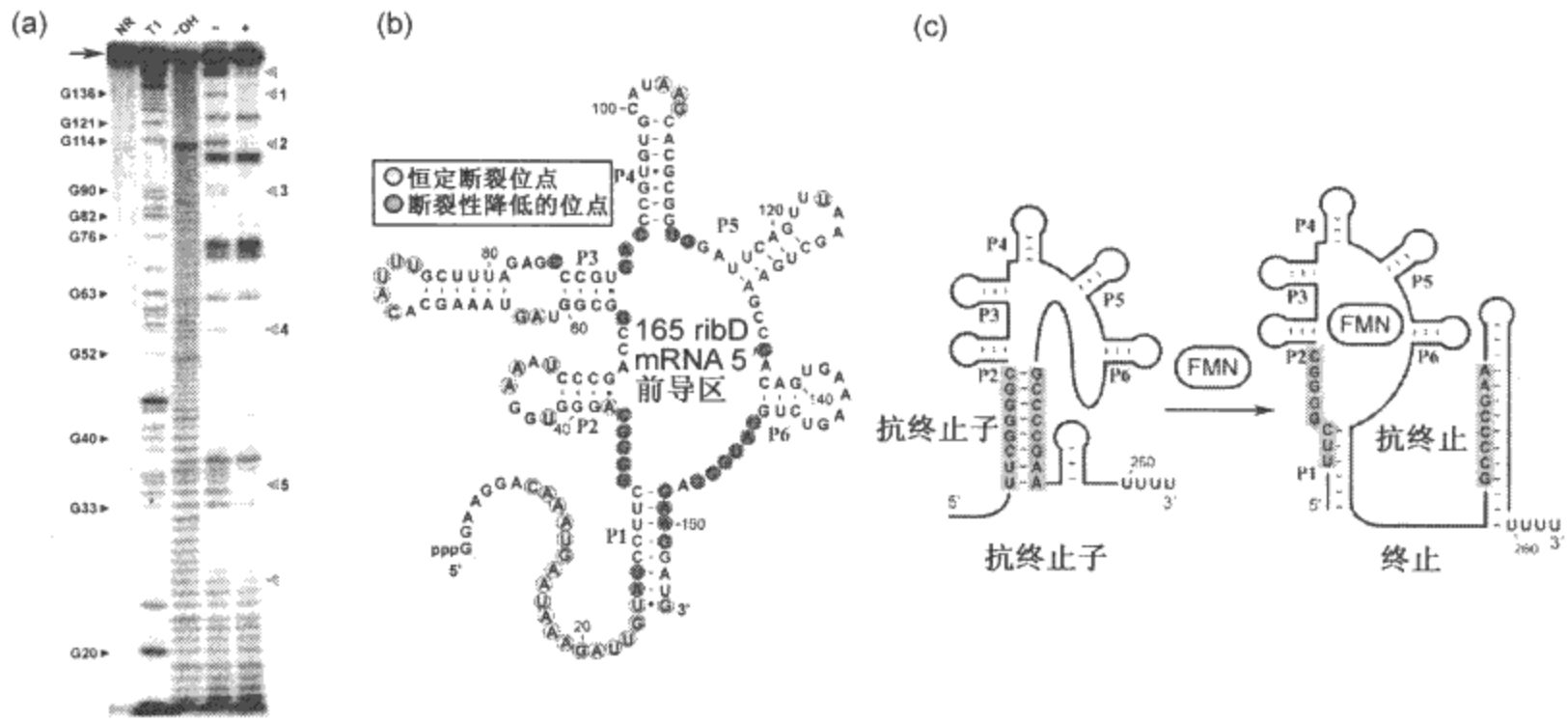


图 7.33　RFN 元件的线内探测结果及 *ribD* 核糖开关的作用模型。（a）线内探测实验的凝胶电泳结果。泳道 1，无 RNA。泳道 2，RNaseT1 切割的 RNA。泳道 3，碱切割的 RNA。泳道 4 和 5，FMN 无（－）和有（＋）条件下 25℃自发水解 40h 的 RNA。右边箭头指示有 FMN 时对剪切不敏感的 RNA 区域。（b）枯草杆菌 *rib* mRNA 的 5′-UTR 的部分序列，显示因 FMN 结合而对自发剪切不敏感的核苷酸之间的连接（红色），以及不受 FMN 影响保持原有敏感性的核苷酸之间的连接（黄色）。元件的二级结构是基于许多 RFN 元件的序列比对而推出的。（c）FMN 结合后核糖开关可能的变化。无 FMN 时，两个黄色区域之间碱基配对，迫使核糖开关采取抗终止构型，其发夹远离 U 串。相反，FMN 与生长的 mRNA 链的结合使 GCCCCGAA 序列与核糖开关的另一部分碱基配对，形成终止子，导致转录终止。（*Source*：(a-c) © 2002 National Academy of Science. Proceedings of the National Academy of Sciences，Vol. 99，no. 25，December 10，2002，p. 15908-15913 “An mRNA struc-ture that controls gene expression by binding FMN,” Chalamish，and Ronald R. Breaker，fig. 1，p. 15909 & fig. 3，p. 15911.）

件和终止子序列的克隆 DNA 为模板，进行体外转录分析实验，发现即使无 FMN，也有 10%的转录在终止子处终止。但 FMN 的存在使终止转录的概率提高到 30%。研究者利用失控转录分析（第 5 章）对终止位点作图，发现转录恰好终止在 U 串的末端。随后，研究者以突变 DNA 为模板进行体外转录，该模板在推测的终止子处编码的 U 少于 6 个。结果 FMN 没有改变转录终止的频率，可能是因为 U 串序列太短，降低了终止子的效率。因此，对野生型基因而言，FMN 确实迫使正在生长的转录物形成终止子结构，导致转录的停顿。

在第 17 章学习翻译调控时将了解核糖开关的另一个事例。配体与 mRNA 5′-UTR 的适体结合后，通过改变 5′-UTR 的构型使核糖体结合位点不能与核糖体结合而调控 mRNA 的翻译。

在这两个事例中，核糖开关的作用都是抑制基因表达，一个是在转录水平上，一个是在翻译水平上。目前，有关核糖开关的研究数据显示，核糖开关确实在这两个水平上抑制基因的表达。至于为什么核糖开关不能促进基因表达，目前还没有合理的解释。基于对多种核糖开关的研究，归纳推导出了核糖开关的一般作用模型（图 7.34）。在 mRNA 的 5′-UTR 存在着两个结构域：适体结构域和表达平台区（expression platform）。表达平台区可以是终止子、核糖体结合位点或其他影响基因表达的 RNA 元件。通过与适体结合并改变核糖开关的构象，配体能够影响表达平台区从而调控基因表达。

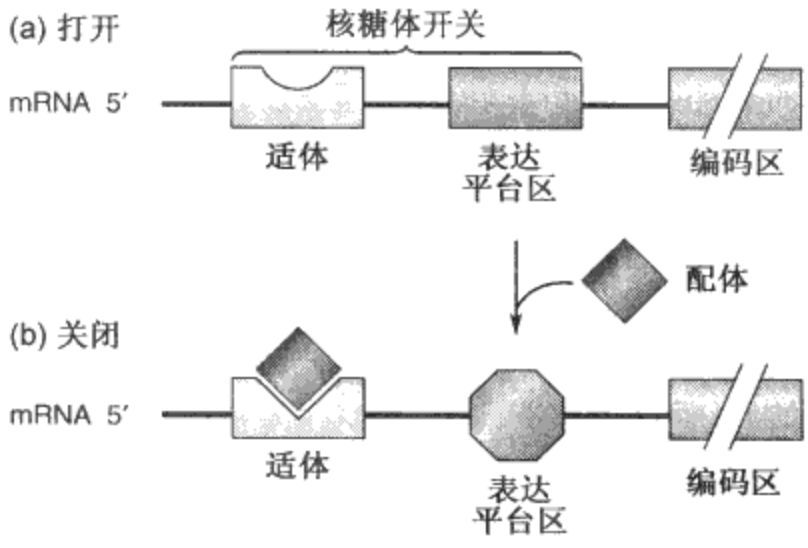


图 7.34　核糖开关作用模型。（a）配体缺乏时，基因处于表达状态。（b）配体存在时，配体与核糖开关的适体结合，引起核糖开关、表达平台区构象改变，从而关闭基因的表达。

核糖开关是变构调控基因表达的另一个事例，即配体引起大分子构象的改变，进而影响了大分子与其物质相互发生作用的能力。在本章中我们学过 *lac* 操纵子的表达调控就是一种变构调控机制，配体（异乳糖）与蛋白质（*lac* 阻遏物）结合，干扰了阻遏物与 *lac* 操纵基因的结合。事实上，变构调控的事例很多，以前一直认为变构调控需要变构蛋白的参与，核糖开关的调控机制与之类似，只是变构的大分子是 RNA 而不是蛋白质。

最后，核糖开关可能为我们开启了一扇认识“RNA 世界”的窗口，“RNA 世界”可能存在于生命进化的早期，此时蛋白质和 DNA 还没有进化形成。在“RNA 世界”中，基因由 RNA 而不是 DNA 构成，酶也是由 RNA 而不是蛋白质组成。（在第 14、17 和 19 章我们将了解具有催化功能的 RNA），无须蛋白质调控基因，RNA 世界的生命依靠小分子与基因直接相互作用。如果这个假说是真实的，那么，核糖开关就是遗传调控最古老形式的遗迹。

小结 位于 mRNA 5′-UTR 的核糖开关包含 2 个结构域：能够与配体结合的适体结构域，以及通过构象改变而引起基因表达变化的表达平台区。例如，FMN 能够结合 RFN 元件，RFN 元件是 *ribD* 操纵子 mRNA 5′-UTR 的核糖开关上的适体。FMN 与适体结合后，核糖开关的碱基配对方式发生改变而产生终止子结构，使转录衰减，这有助于细胞节约能量，因为 FMN 是 *ribD* 操纵子的表达产物之一。

总结

E. coli 细胞对乳糖的分解代谢需要 β-半乳糖苷酶和半乳糖苷透性酶来完成。编码这两种酶以及另外一种酶的基因簇集在一起，从单一启动子处起始转录，形成一条多顺反子信使 RNA。这些功能相关的基因能够被共同调控表达。

lac 操纵子的表达具有正负两种调控机制。其负调控机制如下：由于阻遏物能够阻止 RNA 聚合酶与启动子结合而对 *lac* 基因的转录，所以，一旦阻遏物与操纵基因结合，操纵子即被关闭。当可被利用的葡萄糖被消耗殆尽且有乳糖可被利用时，少数的 *lac* 操纵子酶分子将乳糖转换为异乳糖，异乳糖作为诱导物与阻遏物结合，引起阻遏物构象发生改变，促使阻遏物与操纵基因解离。随着阻遏物的解离，RNA 聚合酶与启动子结合而起始 *lac* 基因的转录。联合遗传学与生物化学实验的研究结果显示，操纵基因和阻遏物基因是 *lac* 操纵子负调控系统的两个关键元件。DNA 测序结果显示，除主操纵基因外，还存在 2 个辅操纵基因，分别位于主操纵基因的上游和下游，这 3 个操纵基因都是高效阻遏所必需的。

lac 操纵子以及其他一些编码参与糖代谢的酶的操纵子，其正调控由代谢激活蛋白（CAP）介导，CAP 与 cAMP 协同促进操纵子的转录。由于葡萄糖的存在会引起 cAMP 浓度的降低，所以葡萄糖能够阻止操纵子的正调控。只有在葡萄糖浓度较低且需要代谢另一种糖源时，*lac* 操纵子才能被激活。CAP-cAMP 复合体通过结合在邻近启动子的激活子结合位点上而促进 *lac* 操纵子的转录，CAP-cAMP 的结合有助于 RNA 聚合酶与启动子的结合。CAP-cAMP 与激活子结合位点结合后，能够招募 RNA 聚合酶与 *lac* 启动子的结合，形成闭合启动子复合体，然后由闭合启动子复合体转换为开放启动子复合体。CAP 通过与 RNA 聚合酶的 αCTD 发生蛋白质-蛋白质间相互作用而招募 RNA 聚合酶与启动子的结合。

ara 操纵子受 AraC 蛋白的调控。AraC 蛋白能够使相距 210bp 的 $araO_2$ 和 $araI_1$ 两位点之间的 DNA 发生环化，从而阻遏 *ara* 操纵子的表达。阿拉伯糖的结合能够松弛 AraC 蛋白与 $araO_2$ 之间的结合，进而与 $araI_1$ 和 $araI_2$ 结合，破坏了形成的 DNA 环，使 *ara* 操纵子去阻遏而发生转录。CAP 和 cAMP 通过与位于 *araI* 上游的一个位点结合而进一步促进了 *ara* 操纵子的转录。AraC 蛋白通过与 $araO_1$ 位点的结合阻止 *ara*C 基因由右向左的转录而调控自身的合成。

trp 操纵子在细胞内色氨酸浓度很高时，能够应答包括辅阻遏物色氨酸在内的阻遏物，辅阻遏物与无辅基阻遏物结合，使之构象发生改变，能够更好地与 *trp* 操纵基因结合，导致操纵子被阻遏。

E. coli trp 操纵子的衰减机制能够在色氨酸充足的情况下发挥作用。当色氨酸的供应趋紧时，核糖体在 *trp* 操纵子前导区的 2 个连续的色氨酸密码子处停顿。由于 *trp* 前导区正处于转录过程中，因此核糖体的停顿会影响 RNA 分子的折叠方式，特别是能够影响携带部分转录终止信号的发夹结构的形成，该转录终止信号能够引发衰减作用。因此，当色氨酸缺乏时，衰减作用失效，操纵子仍然处于激活状态。这意味着，同阻遏作用一样，衰减作用通过感应色氨酸浓度水平而发挥调控作用。

位于 mRNA 5′-UTR 的核糖开关包含 2 个结构域。能够与配体结合的适体结构域，以及通过构象改变而引起基因表达变化的表达平台区。例如，FMN 能够结合 RFN 元件，RFN 元件是 *ribD* 操纵

子 *m*RNA 5′-UTR 的核糖开关的适体。FMN 与适体结合后核糖开关的碱基配对方式改变，产生终止子结构，使转录衰减。

复习题

1. 绘制 *E. coli* 细胞在含有葡萄糖和乳糖培养基上的生长曲线。曲线的每一部分反映了什么事件的发生？

2. 图示说明 *lac* 操纵子的正负调控机制。

3. β-半乳糖苷酶和半乳糖苷透性酶的功能是什么？

4. 为什么 *lac* 操纵子的正负调控机制对于 *E. coli* 细胞的能量效率十分重要？

5. 描述证实 *lac* 操纵基因是阻遏物结合位点的实验方案，并给出实验结果。

6. 描述实验并给出结果，证明即使阻遏物已结合在操纵基因上，RNA 聚合酶也能与 *lac* 启动子结合。

7. 描述实验并给出结果，显示 *lac* 阻遏物阻止 RNA 聚合酶结合 *lac* 启动子。

8. 如何得知 3 个操纵基因是完全阻遏 *lac* 操纵子所必需的？去除 1 个或 2 个辅操纵基因的相对效应是什么？

9. 描述实验并给出结果，显示 cAMP 可在一定水平上促进 β-半乳糖苷酶的合成。使用野生型和突变型细胞提取物，其中突变降低了 CAP 对 cAMP 的亲和力。

10. 提出关于 CAP-cAMP 促进 *lac* 操纵子转录的假说，其中包括 RNA 聚合酶 α 亚基的 C 端结构域（αCTD）。

11. 描述电泳实验并给出结果，证明 CAP-cAMP 的结合引起 *lac* 启动子区域弯曲。

12. 支持 CAP-cAMP 结合引起 DNA 弯曲的其他证据是什么？

13. 在 *araBAD* 操纵子中，$araO_2$ 和 *araI* 位点之间插入整数螺旋转角的 DNA 片段（10.5 的倍数）时，AraC 可阻遏 *araBAD* 操纵子的表达，但插入非整数螺旋转角 DNA 时却阻止了 AraC 的阻遏作用。请解释并图示这一现象。

14. 图示阿拉伯糖如何解除对 *araBAD* 操纵子的阻遏作用。显示（a）无阿拉伯糖和（b）有阿拉伯糖时 AraC 的结合位点。

15. 描述实验并给出结果，证明阿拉伯糖能够破坏 AraC 参与形成的阻遏环。

16. 描述实验并给出结果，证明 $araO_2$ 和 *araI* 均参与阻遏环的形成。

17. 描述实验并给出结果，证明从打开的阻遏环上去除阿拉伯糖后，阻遏环能重新形成。

18. 描述实验并给出结果，证明在开环而非成环状态下 $araI_2$ 在 AraC 结合中是重要的。

19. 提出解释 *E. coli* 细胞 *trp* 操纵子负调控的模型。

20. 提出解释 *E. coli* 细胞 *trp* 操纵子衰减作用的模型。

21. 在 *E. coli* 细胞中，为什么 *trp* 前导区的翻译不是简单地延续到 *trp* 结构基因（如 *trpE* 等）？

22. 色氨酸饥饿时，*E. coli* 细胞如何克服 *trp* 衰减作用？

23. 何谓核糖开关？举例说明。

24. 描述“线内探测”的含义。

分析题

1. 表中给出了几个部分二倍体 *E. coli* 菌株的基因型（就 *lac* 操纵子而言），请填写表型。“+”表示合成 β-半乳糖苷酶，“−”表示不合成 β-半乳糖苷酶。在所有情况下都缺乏葡萄糖，简要说明你的理由。

β-半乳糖苷酶合成的表型		
基因型	有诱导物	无诱导物
a. $I^+O^+Z^+ / I^+O^+Z^+$		
b. $I^+O^+Z^- / I^+O^+Z^+$		
c. $I^-O^+Z^+ / I^+O^+Z^+$		
d. $I^sO^+Z^+ / I^+O^+Z^+$		
e. $I^+O^cZ^+ / I^+O^+Z^+$		
f. $I^+O^cZ^- / I^+O^+Z^+$		
g. $I^sO^cZ^- / I^+O^+Z^+$		

2. （a）在下表列出的基因型中，字母 A、B、C 对应 *lacI*、*lacO*、*lacZ* 基因座，但无须考虑它们的次序。请根据表中列出的前 3 个突变体的表型，推导 A、B、C 所对应的基因座。上角标“−”（如 A^-）表示基因功能异常：Z^-、O^- 和 I^-。

（b）根据常规 *lac* 操纵子遗传表示方法，确定表中第 4 和 5 行列出的部分二倍体菌株的基因型。其中 I^+、I^- 和 I^s 都是可能的。

β-半乳糖苷酶合成的表型		
基因型	有诱导物	无诱导物
1. $A^+B^+C^-$	+	+
2. $A^-B^+C^+$	+	+
3. $A^+B^+C^- / A^+B^+C^+$	+	+
4. $A^+B^+C^- / A^+B^+C^+$	−	−
5. $A^-B^+C^+ / A^+B^+C^+$	−	+

3. 就给定的 *E. coli* 细胞而言，每一细胞都会发生下列其中一种突变之一。

a. *lac* 操纵基因突变体（O^c 基因座），不能结合阻遏物。

b. *lac* 阻遏物突变体（I^- 基因产物），不能结合 *lac* 操纵基因。

c. *lac* 阻遏物突变体（I^s 基因产物），不能结合异乳糖。

d. *lac* 启动子突变体，不能结合 CAP 及 cAMP。

每种突变会对 *lac* 操纵子的功能产生怎样的影响（假定无葡萄糖存在）?

4. 你正在研究一种新的 *E. coli* 操纵子，该操纵子与苯基丙氨酸生物合成有关。

a. 预测该操纵子是如何调控的（被苯基丙氨酸诱导还是阻遏？正调控还是负调控?）。为什么？

b. 对操纵子测序发现，在操纵子的 5′端附近有一短的可读框，含有几个苯基丙氨酸密码子，关于这段前导区序列及其编码的前导肽你做何预测？

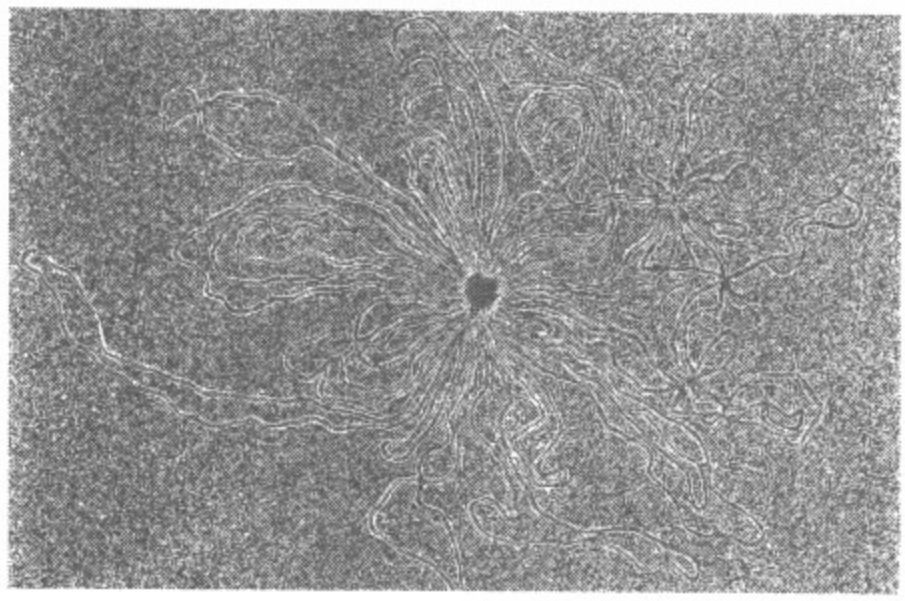
c. 如果改变前导区序列，使苯基丙氨酸密码子（UUU，UUU）突变为亮氨酸密码子（UUA，UUG），结果会怎样？

d. 这是一种怎样的基因表达调控机制？真核生物细胞是否也拥有这种调控机制？其原因是什么？

5. 你怀疑 *E. coli* X 基因的 mRNA 含有适体，且能够与小分子 Y 结合。描述能够验证该假设的实验方案。

翻译　徐启江　校译　李玉花　周　波

第8章　细菌转录机制的主要改变



渗透胁迫处理的 T4 噬菌体释放 DNA 的假彩色透射电镜照片

© *Biozentrum/SPL/Photo Researchers, Inc.*

在第 7 章，我们讨论了细菌同时调控少数几个基因表达的模式。如 *lac* 操纵子开启时，只激活 3 个基因。在细菌生活史的其他时期，会发生基因表达的根本性改变，这需要转录机制的转换，而操纵子模型无法完成这种转换。本章将讨论引起转录改变的 3 种主要机制，即 σ 因子转换、RNA 聚合酶转换和抗终止。以 λ 噬菌体来阐明抗终止机制，并讨论 λ 噬菌体从一种感染策略转变到另一种策略的遗传机制。

8.1　σ 因子的转换

噬菌体感染细菌后，常会将宿主的转录元件据为己有。在此过程中形成时间依赖，或称时序性转录模式。就是说噬菌体先转录早期基因，然后转录后期基因。在 T4 噬菌体对 *E. coli* 的感染后期，宿主基因基本不再转录，只有噬菌体基因的转录。这种转录特异性的巨大变化很难用第 7 章描述的操纵子机制来解释，而是在转录机制中发生了基础性的改变，即 RNA 聚合酶自身发生了变化。

基因表达的另一个重大改变发生在枯草芽孢杆菌等细菌孢子形成时，在这一时期，营养生长阶段所需基因的表达关闭而形成孢子的特异基因开始表达。同样，这种转换也是由 RNA 聚合酶自身变化引起的。当遭遇饥饿、热激、营养缺乏等胁迫时，细菌会通过改变转录模式而做出响应。

因此，细菌对其生活环境改变所做出的响应是通过对转录调控的全面改变而实现的，并且这种转录改变伴随 RNA 聚合酶的改变，更多时候是 σ 因子的改变。

噬菌体感染

RNA 聚合酶的哪一部分最可能调控酶的专一性呢？在第 6 章介绍过 σ 因子是在离体条件下决定噬菌体 T4 DNA 转录特异性的关键因子，因此，σ 因子是最合理的答案，实验证明确实如此。但这些实验并不是首先在 *E. coli* T4 系统进行的，而是在枯草芽孢杆菌及其噬菌体特别是 SPO1 进行的。

与 T4 噬菌体一样，SPO1 也有一个大的基因组。其时序转录模式为：在感染后 5min 左右，早期基因表达；接着，中期基因表达启动（在感染 5～10min 后）；从 10min 后直到感染结束，晚期基因表达启动。由于噬菌体有很多基因，它利用精细的机制来调控转录便不足为奇了。Janice Pero 及同事在这一模式的研究中处于领先地位（图 8.1）。

与 T4 噬菌体的模式相似，宿主 RNA 聚合酶全酶控制 SPO1 早期基因的转录，即最早的基因由宿主全酶转录（第 6 章）。这是必需的，因为噬菌体自身并不携带 RNA 聚合酶。当噬菌体第一次感染细胞时，宿主的全酶是唯一可用的 RNA 聚合酶。枯草芽孢杆菌的全酶与 *E. coli* 的很相似，其核心酶包括两个大亚基（β 和 β′），两个小亚基（α）和一个很小的亚基（ω）。其初级 σ 因子的相对分子质量为

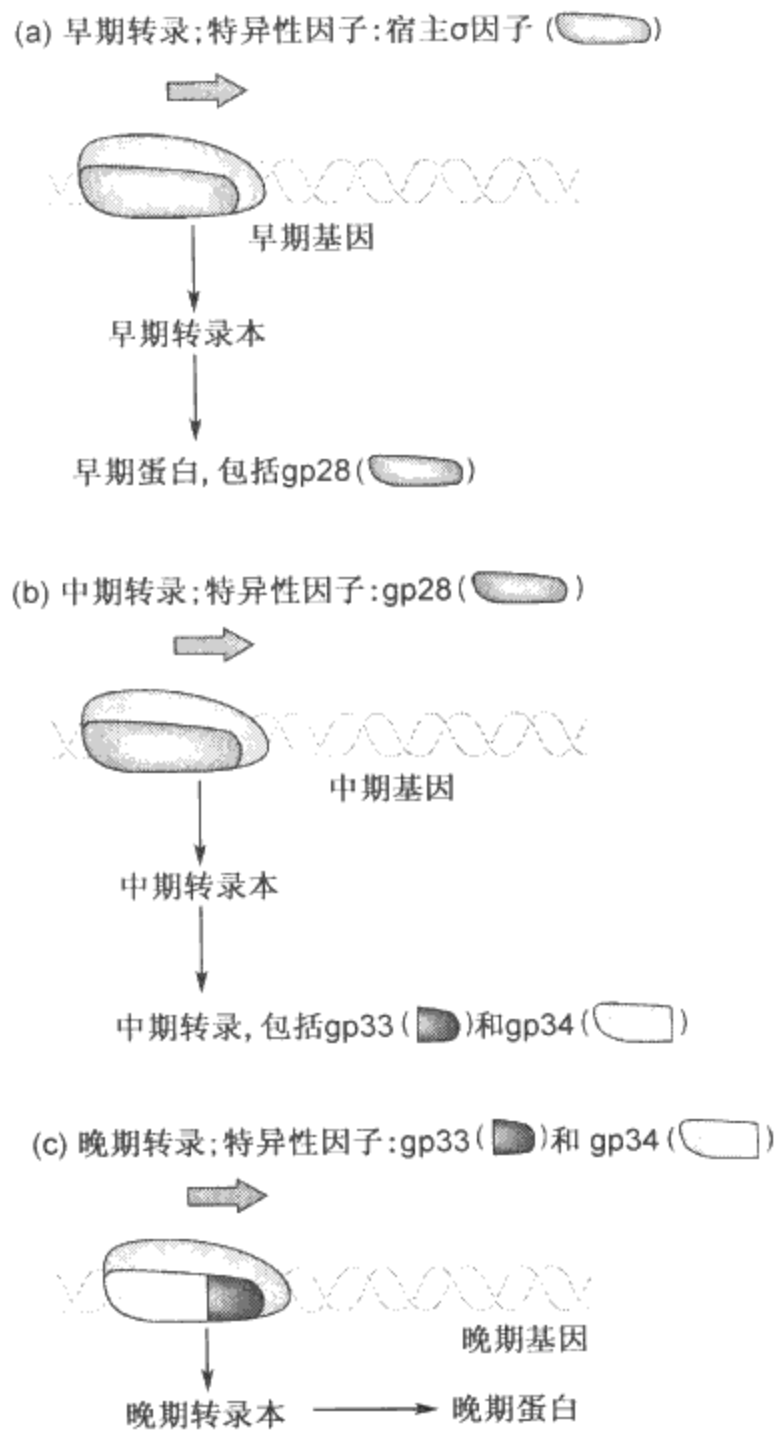


图 8.1　感染枯草芽孢杆菌的 SPO1 噬菌体转录的时序控制。(a) 宿主 RNA 聚合酶全酶，包括宿主 σ 因子（蓝色）指导早期转录。gp28（绿色）是一种新的 σ 因子，属于早期噬菌体蛋白。(b) 中期转录由 gp28 与宿主核心聚合酶（红色）共同控制。gp33 和 gp34 是两个中期噬菌体蛋白，共同组成另一个 σ 因子。(c) 晚期转录由宿主的核心聚合酶与 gp33 和 gp34 共同控制。

43 000，较 *E. coli* 的（70 000）为小；此外，该聚合酶包括一个相对分子质量约 20 000 的 δ 亚基，该亚基有助于防止聚合酶结合到非启动子区域，这一功能在 *E. coli* 中由 σ 因子完成。

在 SPO1 感染的早期转录基因 28，其产物 gp28 与宿主核心聚合酶结合，取代宿主的 σ 因子（σ^{43}）。发生噬菌体多肽取代时，RNA 聚合酶的特异性就发生了改变，不再转录早期基因及宿主基因转而开始转录噬菌体中期基因。换言之，gp28 是个新 σ 因子，使宿主 RNA 聚合酶不再转录宿主基因，同时将早期基因转录转向中期基因转录。

从中期向晚期转录的转换也同样如此，只需两个多肽 gp33 和 gp34 共同与核心聚合酶结合并改变其特异性即可。这两个多肽是两个噬菌体中期基因的产物（分别是基因 33 和基因 34），组成 σ 因子取代 gp28，并指导已改变的聚合酶转录噬菌体晚期基因而非中期基因。注意，在此过程中宿主核心聚合酶始终保持完整，只是通过 σ 因子的不断替换来改变聚合酶的特异性，进而指导转录进程。当然，转录特异性的改变也有赖于早、中、晚期基因启动子序列的差异，正因为如此，它们才能被不同 σ 因子所识别。

这一过程的最大特点是不同 σ 因子的大小不同。宿主 σ 因子、gp28、gp33 和 gp34 的相对分子质量分别为 43 000、26 000、13 000 和 24 000，但它们都可以与核心聚合酶结合并行使 σ 因子的功能（gp33 和 gp34 必须共同行使这一功能）。事实上，*E. coli* 中相对分子质量为 70 000 的 σ 因子在离体实验中也可与枯草芽孢杆菌的核心聚合酶结合，也说明核心聚合酶具有可变的 σ 因子结合位点。

如何确定 σ 因子转换模式的有效性呢？遗传学和生物化学证据都支持这一模式。首先，遗传学研究表明，基因 28 突变可抑制从早期基因向中期基因转录的转换，证明基因 28 产物是开启中期基因转录的 σ 因子。同样，基因 33 和 34 的突变也抑制了从中期基因向晚期基因转录的转换，亦与此模式相符。

Pero 及同事进行了生物化学研究。首先，他们从 SPO1 感染的细胞中纯化出 RNA 聚合酶。纯化过程包括一个磷酸纤维素层析分离步骤，可分离聚合酶的 3 种形态。首先分离出的聚合酶 A 包含宿主核心酶、δ 亚基及所有噬菌体编码的因子，其他两个聚合酶，B 和 C，缺少 δ 亚基，而 B 中包括 gp28，C 中包括 gp33 和 gp34。图 8.2 展示了通过 SDS-PAGE 检测到的后面两个聚合酶的亚基结构。由于缺少 δ 亚基，这两种酶不能清楚地识别 DNA 的启动子及非启动子区域，因此无法进行特异性转录。而当 Pero 及同事将 δ 亚基加入聚合酶并检测其转录特异性时，发现聚合酶 B 特异性地转录噬菌体的晚早期基因，而聚合酶 C 则特异性地转录噬菌体晚期基因。

小结　噬菌体 SPO1 感染的枯草芽孢杆菌，其基因转录以时序模式进行：早期基因先转录，然后是中期基因，最后是晚期基因。这种转换由噬菌体编码的一系列 σ 因子来调控。这些 σ 因子与宿主核心聚合酶结合从而改变聚合酶对早、中、晚期基因启动子识别的特异性。宿主 σ 因子指导噬菌体早期基因转录，噬菌体蛋白 gp28 将转录特异性转换为中期基因，而噬菌体蛋白 gp33 和 gp34 再将其转换为晚期基因。

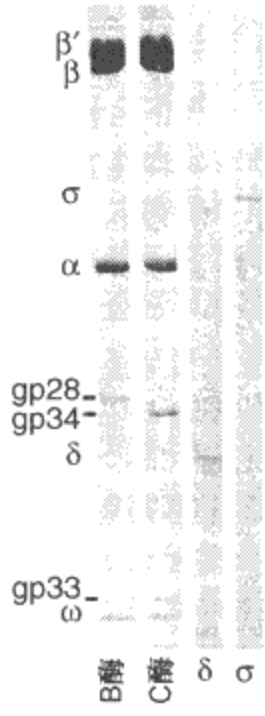


图 8.2　噬菌体 SPO1 感染的枯草芽孢杆菌细胞中 RNA 聚合酶的亚基组成。层析分离 RNA 聚合酶，然后经 SDS-PAGE 显示其亚基组成。B 酶（泳道 1）含有核心亚基（β′、β、α 和 ω）和亚基 gp28。C 酶（泳道 2）含有核心亚基 gp34 和 gp33。最后两道分别含有 δ 亚基和 σ 亚基。（*Source*：Pero J.，R. Tjian，J. Nelson，and R. Losick. *In vitro* transcription of a late class of phage SPO1 genes. *Nature 257*（18 Sept 1975）：f. 1，p. 249 © Macmillan Magazines Ltd.）

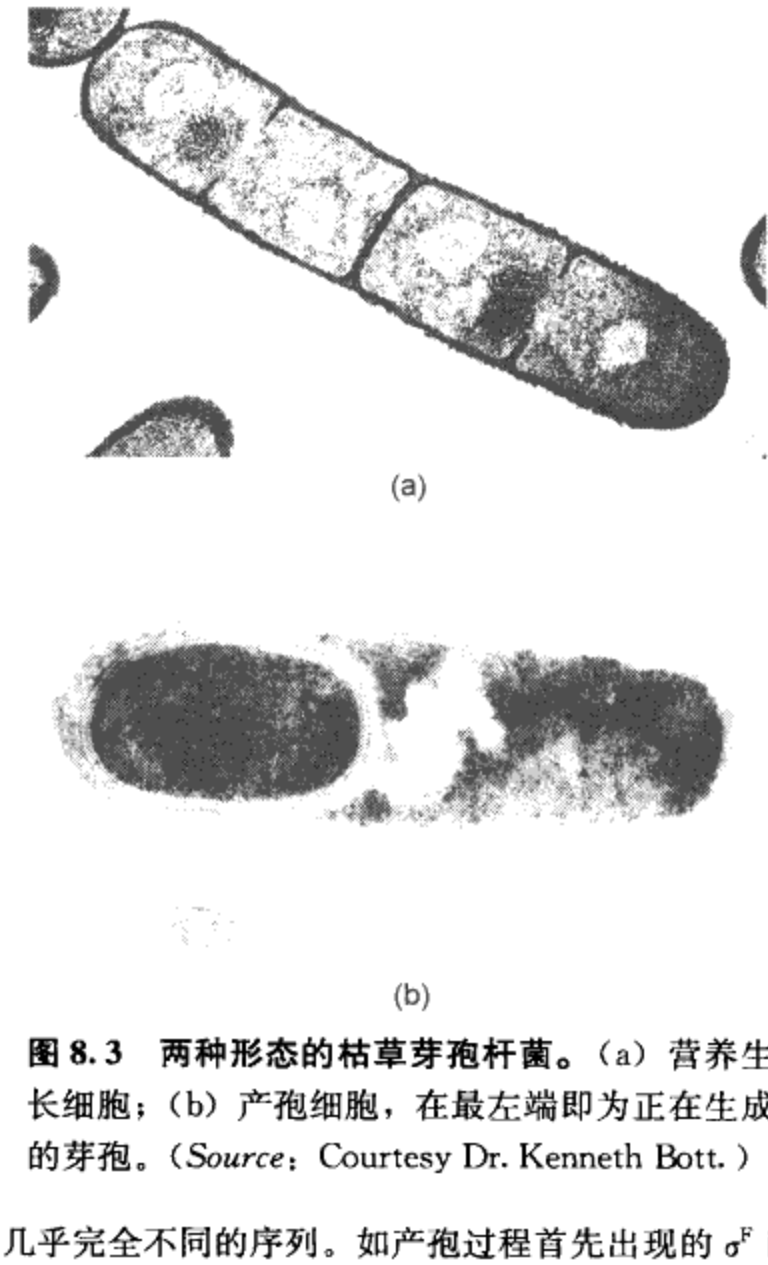


图 8.3　两种形态的枯草芽孢杆菌。（a）营养生长细胞；（b）产孢细胞，在最左端即为正在生成的芽孢。（*Source*：Courtesy Dr. Kenneth Bott.）

孢子形成

以上讨论了 SPO1 噬菌体如何通过 σ 因子替换来改变宿主 RNA 聚合酶的特异性。现在，我们讨论在宿主孢子形成（sporulation）过程中基因表达发生改变的相同机制。枯草芽孢杆菌在养分和其他条件适合的营养生长环境（vegetative）中可无限繁殖，但是在饥饿或其他不利条件下，便会形成坚硬的芽孢（endospore），在休眠状态中芽孢可以存活多年，直到条件适宜才重新生长（图 8.3）。

在形成芽孢过程中，基因表达肯定会发生变化，因为不同形态和代谢过程中营养细胞和芽孢细胞中的某些基因产物不同。事实上，当枯草芽孢杆菌细胞产生芽孢时，会激活一整套新的产孢特异性基因。从营养生长到芽孢形成的转变是通过一个复杂的 σ 转换系统，关闭一些营养生长基因的转录，同时又开启产孢特异性基因的转录。

孢子形成是细菌生长过程的一个基本的变化，涉及大量营养基因的关闭和产孢基因的开启。此外，转录水平上的改变不是绝对的，如在营养生长状态下非常活跃的一些基因，在形成孢子的过程中也保持活性。转录开关机制怎样应付这一过程的复杂呢？

你可能会预计有 1 个以上的新 σ 因子参与孢子形成。事实上，参与此过程的 σ 因子有 σ^F、σ^E、σ^H、σ^G 和 σ^K，另外还包括营养生长期的 σ^A，每个 σ 因子都有识别不同启动子的重要作用。例如，营养生长相关的 σ^A 识别的启动子与 *E. coli* σ 因子所识别的启动子非常相似，−10 框一般为 TATAAT，−35 框的保守序列为 TTGACA，而产孢特异性 σ 因子识别几乎完全不同的序列。如产孢过程首先出现的 σ^F 因子激活编码其他产孢特异性 σ 因子基因的转录。

为了阐明用于证明 σ 因子真实存在所使用的技术，我们先了解一下由 Richard Losick 及同事对其中一个 σ 因子 σ^E 所做的工作。首先，他们证明 σ^E 赋予聚合酶对一个已知产孢基因的特异性。他们用包含 σ^E 或 σ^A 的聚合酶在体外有标记核苷酸条件下，转录含有部分枯草芽孢杆菌 DNA 的质粒。这段枯草芽孢杆菌 DNA（图 8.4）含有营养生长和孢子发生相关基因的启动子。营养生长启动子位于 3050bp 的限制性片段中，孢子发生启动子在 770bp 的限制性片段中。Losick 及同事利用标记的 RNA 产物与模板 DNA 进行 Southern 印迹杂交，显示了 σ 因子的特异性：如果转录营养生长基因，所得标记的 RNA 会在模板 DNA 的 3050bp 处显示 Southern 杂交带。如果转录孢子发生基因，则标记的 RNA 产物与模板 DNA 的杂交带出现在 770bp 处。图 8.5 表明含 σ^A 的聚合酶所转录的 RNA 只出现与营养生长相关的杂交带（3050bp）。而含 σ^E 的聚合酶所转录的 RNA，在营养生长和孢子发生相关基因处（3050bp 和 770bp）均出现杂交信号。显然 σ^E 可识别营养生长基因的启动子，但是主要亲和性是针对孢子发生基因的启动子，至少是对 770bp DNA 片段中的那些类型。

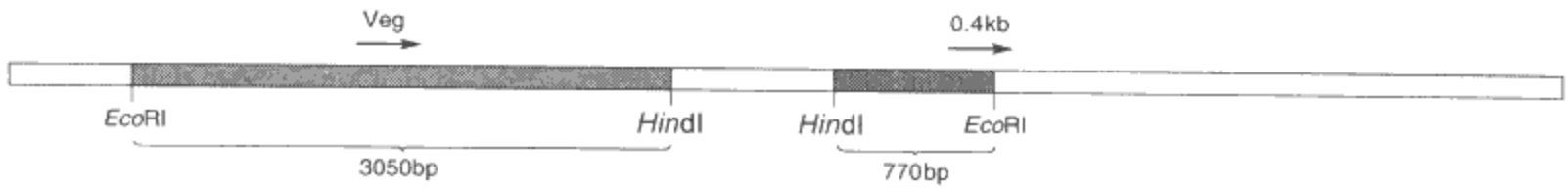


图 8.4　质粒 p213 的部分图示。这段 DNA 含有两个启动子：营养生长启动子（Veg）和孢子发生启动子（0.4kb）。前者位于 3050bp 的 *Eco*RI-*Hin*dI 片段内（蓝色），后者在 770bp 的片段上（红色）。（*Source*：Adapted from Haldenwang W. G.，N. Lang，and R. Losick，A sporulation-induced sigma-like regulatory protein from *B. subtilis*. *Cell* 23：616，1981.）

图 8.5　σ^A 和 σ^E 的特异性。Losick 小组在体外用含有 σ^A（泳道 1）或 σ^E（泳道 1）的聚合酶转录质粒 p213。然后用标记的转录物与含有 *Eco*RI-*Hin*dI 的质粒片段进行 Southern 印迹杂交。如图 8.4 所示，该质粒在 3050bp 的 *Eco*RI-*Hin*dI 片段中有一个营养生长基因的启动子，而在 770bp 片段中有一个孢子发生基因的启动子。因此，营养生长基因的转录物与 3050bp 的片段杂交，而孢子发生基因的转录物则与 770bp 的片段杂交。本图的放射自显影结果显示 σ^A 酶只转录营养生长基因，而 σ^E 酶既可转录营养生长基因又可转录孢子发生基因。（*Source*：Haldenwang W. G.，N. Lang，and R. Losick，A sporulation-induced sigma-like regulatory protein from *B. subtilis*. *Cell* 23 (Feb 1981)，f. 4，p. 618. Reprinted by permission of Elsevier Science.）

由于 770bp 片段所包含孢子发生基因的性质并未证实，Abraham Sonenshein 小组试图证明 σ^E 能够转录一个已知孢子发生基因。他们以孢子发生所需要且已克隆的基因 *spoIID* 进行转录，用分别含有 σ^B、σ^C、σ^E 的 3 种不同聚合酶转录该基因的截短片段，以便产生截断转录物（第 5 章）。之前，体内 RNA 的 S1 作图已证明 *spoIID* 的天然转录起点。由于截短是在天然转录起点下游 700bp 处，因此从正确转录起点开始的体外转录产生一个 700bp 的截断转录物。如图 8.6 所示，只有 σ^E 可产生该转录物，其他 σ 因子都不能使 RNA 聚合酶识别 *spoIID* 启动子。相似的实验证明 σ^A 也不能识别该启动子。

Losick 小组证实 σ^E 本身就是孢子发生基因的产物，最初称为 *spoIIG*。可以推测，该基因的突变会在早期阻止孢子的形成。如果没有 σ 因子识别孢子发生基因，那么这些基因就不会表达，孢子也就无法形成。

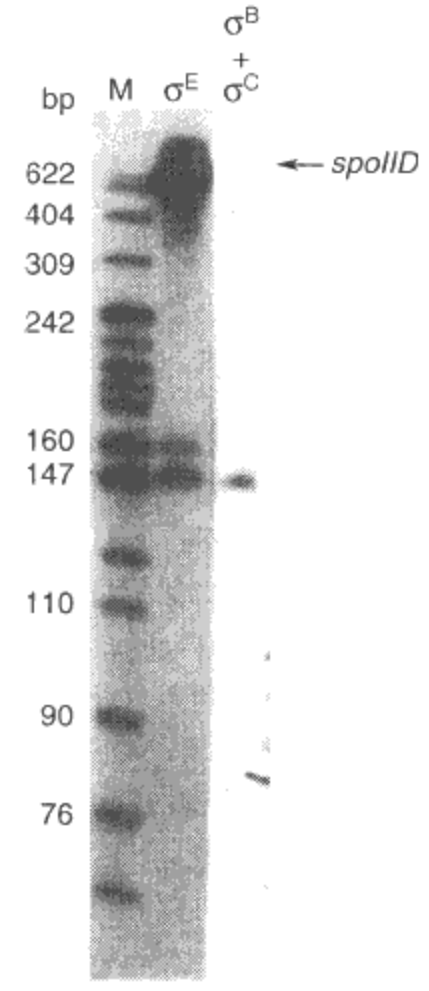


图 8.6　通过 *spoIID* 启动子的截断转录确定 σ^E 的特异性。制备含有 *spoIID* 启动子的限制性酶切片段，然后在体外用枯草芽孢杆菌的核心 RNA 聚合酶加 σ^E（中间泳道），或加 $\sigma^B+\sigma^C$（右道）来转录该片段。泳道 M 是已知各条带大小的 DNA 标准。右边箭头指示预计 *spoIID* 启动子截断转录物的位置（约 770nt），只有含 σ^E 的酶可形成这一转录物。（*Source*：Rong S.，M. S. Rosenkrantz，and A. L. Sononshein，Transcriptional control of the *Bacillus subtilis spoIID* gene. *Journal of Bacteriology* 165，no. 3（1986）f. 7，p. 777，by permission of American Society for Microbiology.）

小结　枯草芽孢杆菌形成孢子时，一组全新的孢子发生特异性基因被开启，许多营养生长基因（但不是全部）被关闭。这种转换主要发生在

转录水平，通过几个新的 σ 因子取代 RNA 聚合酶核心酶中转录营养生长基因的 σ 因子，使孢子发生基因的转录替代营养生长基因的转录，并且每个 σ 因子都有自己偏爱的启动子序列。

拥有多启动子的基因

枯草芽孢杆菌孢子发生的故事对我们的下一个话题多启动子（multiple promoter）是个很好的引言，因为孢子发生基因首先提供了这种现象的实例。当不同 σ 因子占主导地位时，有些基因在孢子发生过程的两个或更多阶段中都要表达，因此这些基因应该有被不同 σ 因子所识别的多个启动子。

spoVG 是具有两个启动子的孢子发生基因，可被 $E\sigma^B$ 和 $E\sigma^E$（分别为包含 σ^B 和 σ^E 的 RNA 聚合酶全酶）所转录。Losick 小组利用 DNA-纤维素层析法从孢子发生细胞中得到全酶的部分分离组分，然后用具有聚合酶活性的分离峰的不同组分对截短的 *spoVG* 基因的克隆进行截断转录。分离峰前缘（leading edge）组分主要产生 110nt 的截断转录物，分离峰后缘（trailing edge）组分主要产生 120nt 的截断转录物，而位于峰中部的组分可同时产生以上两种截断转录物。

利用另一轮 DNA-纤维素层析，他们成功地分离了聚合酶的这两个活性组分。其中，含 σ^E 的组分只能合成 110nt 的截断转录物。此外，产生这一转录本的能力与酶中 σ^E 的含量平行，说明 σ^E 负责这一转录活性。为进一步证明，Losick 小组又用凝胶电泳纯化了 σ^E 并将其与核心聚合酶结合，结果表明只产生 110nt 截断转录物（图 8.7）。同样的实验也证明 σ^B 与核心聚合酶结合后只产生 120nt 的截断转录物。

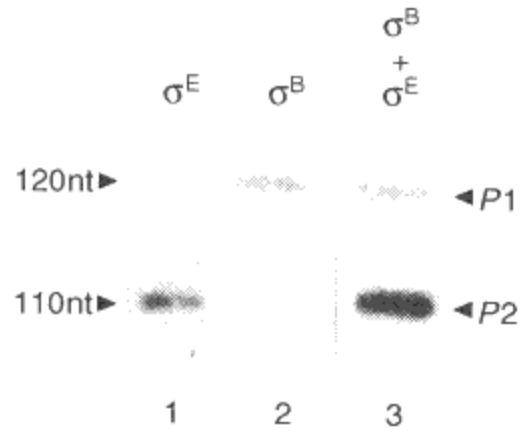


图 8.7　σ^B 和 σ^E 的专一性。Losick 小组通过凝胶电泳纯化了 σ^B 和 σ^E，然后用核心聚合酶通过截断转录进行检测。泳道 1 含 σ^E 引起下游启动子（P2）选择性起始。泳道 2 含 σ^B 引起上游启动子（P1）选择性起始。泳道 3 含有 σ^B 和 σ^E 引起两个启动子的起始。（*Source*: Adapted from Johnson W. C., C. P. Moran, Jr., and R. Losick, Two RNA polymerase sigma factors from *Bacillus subtilis* discriminate between overlapping promoters for a developmentally regulated gene. *Nature* 302 (28 Apr 1983), f. 4, p. 803. © Macmillan Magazines Ltd.)

以上这些实验证明 *spoVG* 基因可被 $E\sigma^B$ 和 $E\sigma^E$ 所转录，而且这两个酶识别的转录起始位点相隔 10bp 的距离，如图 8.8 所示。通过这些起始位点，可以计算其上游的碱基对数，并可以找出每种 σ 因子所识别的启动子区的－10 框和－35 框（图 8.8 所示）。通过与多个被同一 σ 因子所识别的－10 框和－35 框序列的比较，可以确定第 6 章中所报道的保守序列。

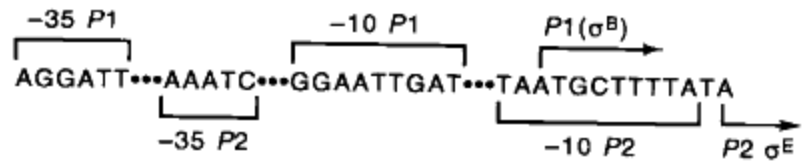


图 8.8　枯草芽孢杆菌 *spoVG* 的重叠启动子。P1 表示由 σ^B 识别的上游启动子，上方用红色表示其转录起点和启动子的－10 框及－35 框。P2 表示由 σ^E 识别的下游启动子，下方用蓝色表示其转录起点和启动子的－10 框及－35框。

小结　有些原核基因在两个不同 σ 因子都有活性时必须被转录，这些基因具有两个启动子，每个启动子由其中一个 σ 因子所识别，从而保证不管哪个 σ 因子存在基因都能表达，同时还能实现在不同情况下基因表达的差异调控。

σ 因子的其他转换

细胞受到高温或其他环境胁迫时会产生**热激反应**（heat shock response），从而使损害降低到最小限度。此时先产生叫做**分子伴侣**（molecular chaperone）的蛋白质，分子伴侣与受到热激而不完全折叠的蛋白质结合，帮助其重新正确折叠。还产生蛋白酶降解那些经分子伴侣帮助仍不能正确折叠的蛋白质。总之，将能编码帮助细胞在热激条件下生存的蛋白质的基因称为**热激基因**（heat shock gene）。

当 *E. coli* 细胞受到热激后（从正常生长温度 37℃升到 42℃），正常转录立即停止或降低，同时有 17 种新的热激蛋白转录物开始合成。这些转录物编码分子伴侣和蛋白酶来协助细胞度过热激胁迫。这一转录过程的转换需要 *rpoH* 基因产物的参与，该基因编码分子质量为 32kDa 的 σ 因子，称为 σ^{32}，也叫 σ^H，这里的 H 代表热激。1984 年 Grossman 小组证明 σ^H 确实是一个 σ 因子，σ^H 与核心聚合酶结合的复合物可以从天然转录起始位点开始，对不同热激基因进行体外转录。

热激反应在 1min 之内就能开始，显然这样短

的时间不足以令 *rpoH* 基因转录并利用 mRNA 翻译出足够多的新 σ 因子。有两个过程可解释 σ^H 是如何迅速积累的。首先，该蛋白本身在温度升高过程中变得更稳定，在正常生长条件下 σ^H 与热激蛋白结合而变得不稳定，发生降解。当温度升高时，许多蛋白质不能正确折叠，使热激蛋白与 σ^H 解离以便拯救或降解那些异常折叠的蛋白质。

其次，高温对 σ^H 浓度的影响反映在翻译水平上：高温引起 *rpoH* mRNA 的 5′-非翻译区的二级结构解离，使其更容易与核糖体结合。Miyo Morita 小组通过对推测的关键二级结构区的突变验证了这一假设。他们发现野生型和突变型 mRNA 二级结构的解离温度与 σ^{32} 合成的可诱导性有关。在第 17 章将详细讨论这一机制。

在氮饥饿情况下，另一 σ 因子（σ^{54} 或 σ^N）直接参与氮代谢相关基因的转录。此外，虽然革兰氏阴性菌，如 *E. coli* 不产生孢子，但对饥饿胁迫也有适应性反应。处于生长稳定期（细胞非增殖期）的 *E. coli*细胞，其 RNA 聚合酶的 σ 因子一旦被 σ^S 替代后，便会使逆境抗性基因的表达开启。这些都是基本的应对机制的例子，细菌往往通过 σ 因子转换所介导的整体转录机制的改变来应对其生存环境的变化。

小结 *E. coli* 的热激反应是由可变的 σ 因子 σ^{32}（σ^H）所控制。σ^{32}（σ^H）取代 σ^{70}（σ^A）后，RNA 聚合酶与热激蛋白基因启动子结合。在高温下 σ^H 的积累是由于 σ^H 稳定性的增加以及其 mRNA 翻译增强所致。而对于低氮或饥饿胁迫下的响应主要依赖于 σ^{54}（σ^N）以及 σ^S。

8.2 T7 噬菌体中所编码的 RNA 聚合酶

T7 噬菌体属于相对简单的一类 *E. coli* 噬菌体，T3 和 ΦII 也属此类。与 SPO1 相比，它们的基因组较小，基因数较少。我们将这类噬菌体的基因转录分为 3 个时期：I、II、III（也可以称为早、中、晚期，以便与 SPO1 中的命名一致）。在 5 个 I 型基因中有一个基因（基因 1）对于 II、III 型基因的表达是必须的。当基因 1 突变后，只有 I 型基因被转录。通过前面对 SPO1 的了解，你可能会认为基因 1 编码一个 σ 因子来指导宿主 RNA 聚合酶转录噬菌体的晚期基因。有人在研究 T7 噬菌体时就得出了这一结论，但这是错误的。

基因 1 的产物不是 σ 因子而是包含在一条多肽里的噬菌体特异性 RNA 聚合酶。该聚合酶特异性地转录 T7 噬菌体的 II 型和 III 型基因，而不转录 I 型基因。该酶确实极具特异性，只转录 T7 噬菌体的 II、III 型基因，而不转录其他任何天然模板。在这种噬菌体中，基因转录的转换机制也相对较简单（图 8.9）。

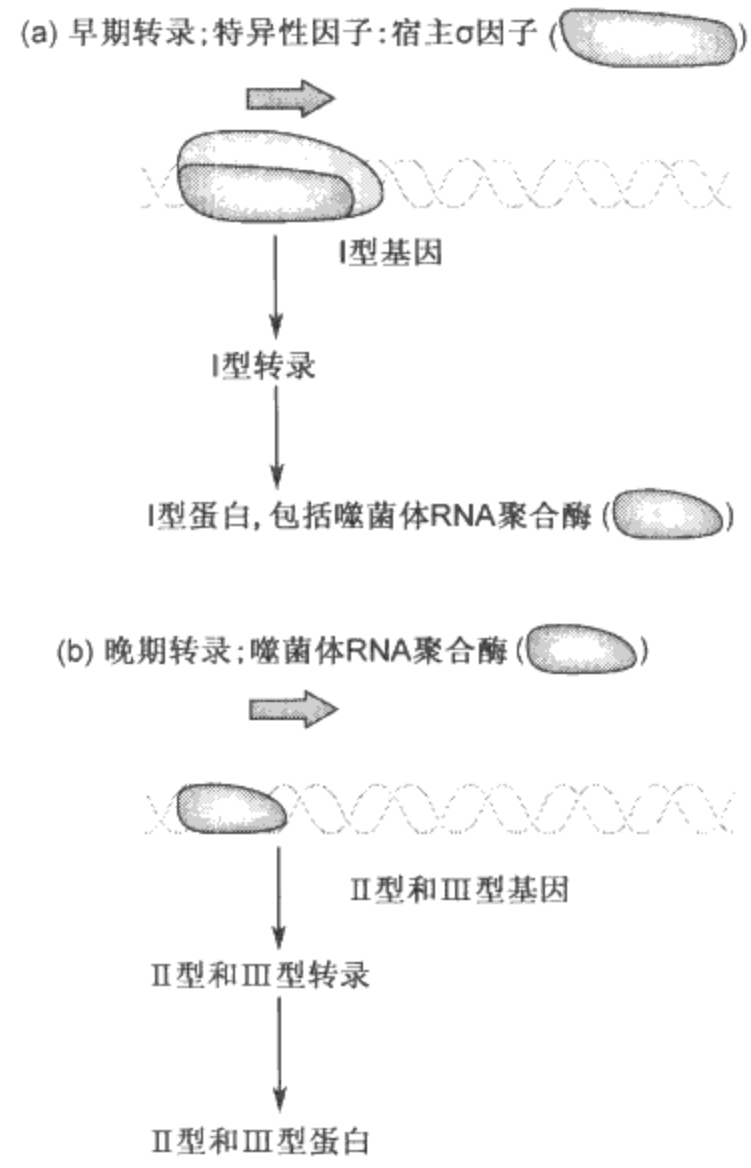


图 8.9 感染 *E. coli* 的 T7 噬菌体转录的时序控制。（a）早期（I 型）转录取决于宿主 RNA 聚合酶全酶，包括宿主 σ 因子（蓝色），其中一种早期噬菌体蛋白是 T7 RNA 聚合酶（绿色）。（b）晚期（II 型和 III 型）转录依赖于 T7 RNA 聚合酶。

当噬菌体 DNA 进入宿主细胞后，*E. coli* 全酶就开始转录噬菌体的 5 个 I 型基因，包括基因 1。基因 1 的产物是噬菌体特异的 RNA 聚合酶，随后它开始转录噬菌体的 II、III 型基因。

从 T3 噬菌体中也已分离到相似的聚合酶。它仅对 T3 基因具有特异性。实际上 T7 和 T3 的启动子已经被构建到克隆载体上，如 pBluescript 载体（见第 4 章）。这些 DNA 在体外可以被其中一个或几个噬菌体聚合酶所转录，产生特异的 RNA 链。

小结 T7 噬菌体编码一个新的对噬菌体晚期基因具有绝对特异性的 RNA 聚合酶，而不是编码新的 σ 因子去改变宿主聚合酶从早期到晚期的特异性。这个由一条多肽组成的聚合酶是噬菌

体最早表达的基因之一（基因 1）的产物。这种噬菌体感染细胞后其转录过程较为简单，宿主聚合酶转录早期基因（I 型基因），其中一个产物是噬菌体聚合酶，该酶转录晚期基因（II 型和 III 型基因）。

8.3 λ 噬菌体感染 *E. coli*

到目前为止，我们所研究的噬菌体（如 T2、T4、T7、SPO1）多数为毒性噬菌体。它们复制时会通过裂解杀死宿主细胞。而 λ 噬菌体是一种**温和**（temperate）噬菌体，感染 *E. coli* 时不一定非要杀死宿主。在这方面 λ 噬菌体比其他噬菌体的感染方式更灵活些，它有两种增殖方式（图 8.10）。第一种是**裂解模式**（lytic mode），与毒性噬菌体感染方式相同。噬菌体 DNA 进入宿主细胞，利用宿主的 RNA 聚合酶进行转录。噬菌体 mRNA 被翻译成子代噬菌体蛋白，噬菌体 DNA 在宿主体内复制并与翻译的蛋白组装成子代噬菌体。当宿主细胞裂解释放子代噬菌体时，感染过程就结束了。

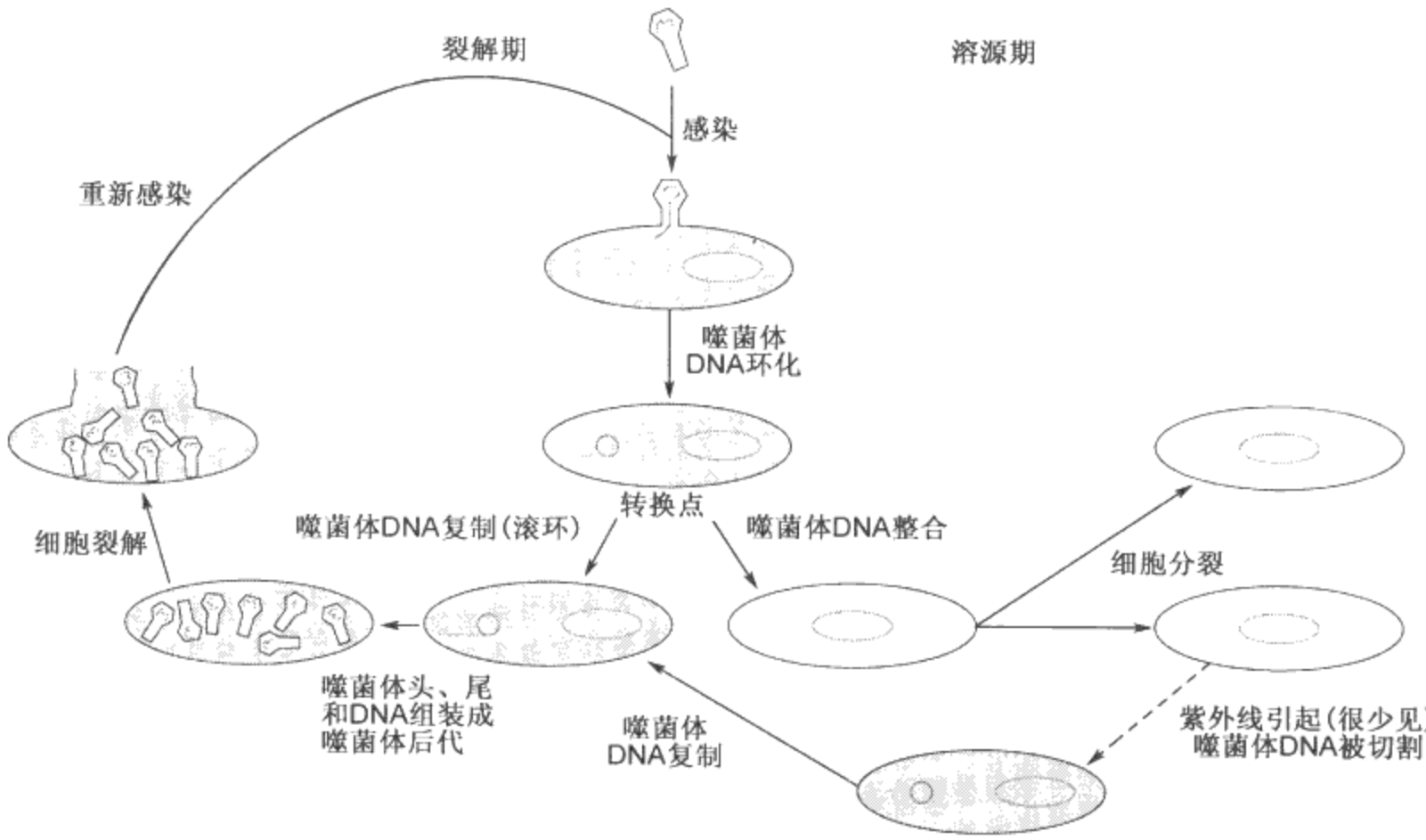


图 8.10　λ 噬菌体引起的裂解性与溶源性感染。蓝色细胞处于裂解期，黄色细胞处于溶源期，绿色细胞尚不确定。

在**溶源模式**（lysogenic mode）中很多过程与裂解不同。噬菌体 DNA 进入宿主细胞后，其早期基因转录并翻译，这一过程与裂解感染一致。但是当 27kDa 的噬菌体蛋白（λ 阻遏物或 CI）出现并与噬菌体的两个操纵子区结合后，所有基因转录都被关闭，只剩下 *cI*（读音“c-one”而不是“c-eye”）这一编码 λ 自身阻遏物的基因。在这种情况下，仅有一个噬菌体基因有活性，这也就很容易理解为什么它没有子代噬菌体形成。而且，选择溶源途径后，噬菌体 DNA 会整合到宿主基因组中。整合了噬菌体 DNA 的细菌也被称为溶源菌（lysogen）。整合的 DNA 被称为**原噬菌体**（prophage）。这种溶源态可以无限期地存在，不能认为这对噬菌体是不利的，因为溶源态噬菌体的 DNA 会随着宿主 DNA 的复制而复制，不用产生噬菌体颗粒就可以成倍地扩增其基因组。因此，可以说它搭上了宿主的“便车”。在特定条件下，如遇到化学诱变剂或辐射，溶源菌便会破裂使噬菌体进入裂解状态。

小结　λ 噬菌体可以通过裂解或溶源两种方式中的任一种进行复制。在裂解模式下几乎所有噬菌体基因都转录并且翻译，同时噬菌体 DNA 复制形成子代噬菌体，然后裂解宿主细胞。在溶源模式下，λDNA 与宿主基因组整合，仅有一个基因表达 λ 阻遏物，λ 阻遏物抑制了噬菌体其他所有基因的转录。然而，整合的噬菌体 DNA 依然能够复制，因为它已成为宿主 DNA 的一部分。

λ噬菌体的裂解繁殖

λ噬菌体的裂解繁殖周期与毒性噬菌体类似，其转录分为3个时期，称为早期（immediate early）、晚早期（delayed early）和晚期（late）。这三类基因依次排列在噬菌体DNA上，这有助于解释它们是如何被调控的。图8.11以两种形式展示了λ噬菌体的遗传图谱。噬菌体处于颗粒状态时DNA为线形，噬菌体感染宿主后短时间内形成环形。环化过程可能是由于线性基因组两端有12bp的黏性末端所致，这种黏性末端称为cos位点。环化作用将分别位于线性基因组两端的晚期基因连接到一起。

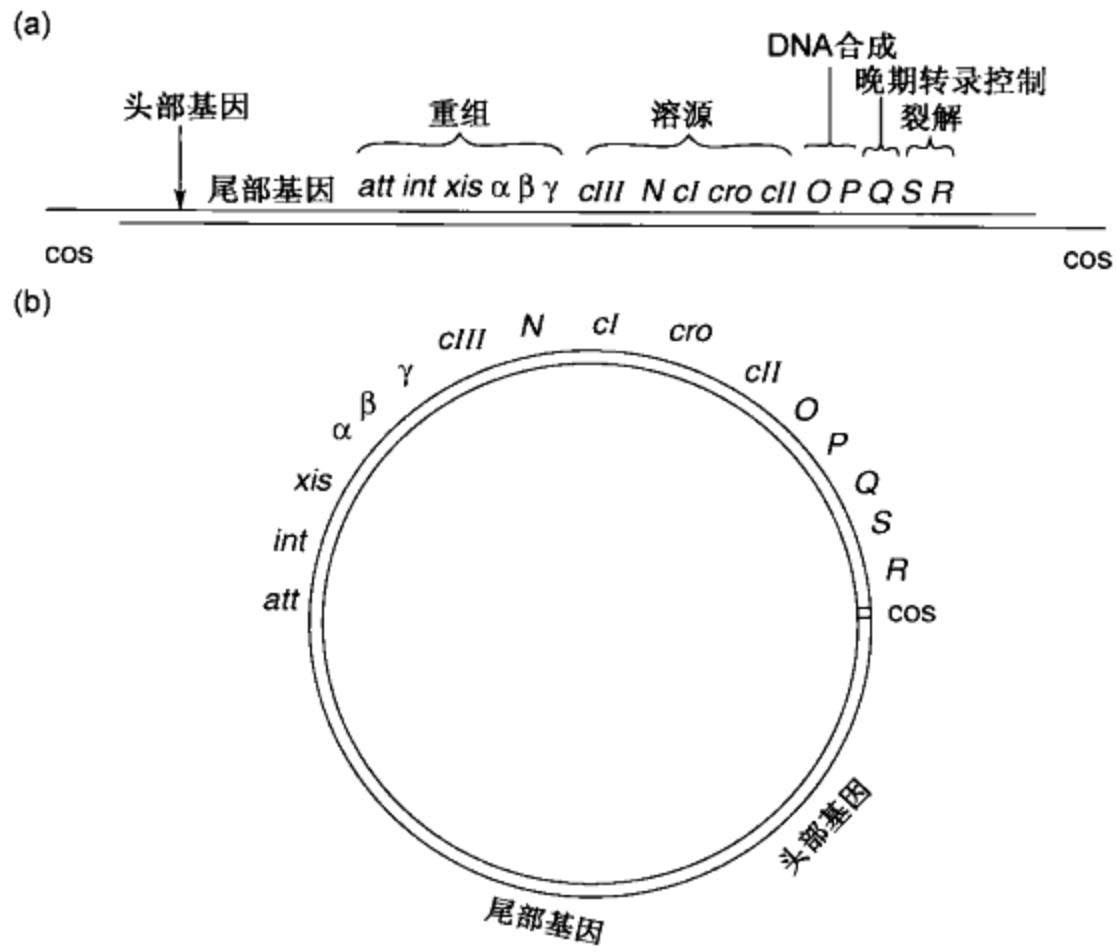


图8.11 λ噬菌体的遗传图谱。（a）线形图，DNA以线性方式存在于λ噬菌体颗粒中，黏性末端（cos）在图的两端。基因基本上是按功能分组。（b）环形图，DNA存在于被裂解性感染的宿主细胞，其黏性末端已连接。

通常，噬菌体基因表达的程序是由转录开关所控制的，但λ噬菌体采用了一种我们还未见过的转换方式：**抗终止（antitermination）**。图8.12是这一过程的示意图。宿主RNA聚合酶全酶先转录噬菌体的早期基因。这类基因只有*cro*和*N*两个，分别直接位于右向启动子P_R和左向启动子P_L下游。在裂解周期的这一阶段，尚无阻遏物与控制这些启动子的操纵子（分别是O_R和O_L）结合，所以转录进程不受到阻遏。当聚合酶到达早期基因的末端时，遇到终止子并在晚早期基因处终止转录。

这两个早期基因产物对λ噬菌体基因的后续表达十分关键。*cro*基因产物是一个阻遏物，能抑制λ阻遏物基因*cI*的转录，阻止λ阻遏物蛋白合成。这一点对受λ阻遏物抑制的其他噬菌体基因的表达十分必要。*N*基因的产物**N蛋白**是个**抗终止因子**（antiterminator），使RNA聚合酶忽略早期基因末端的终止子而继续转录后续的晚早期基因。请注意，早期和晚早期基因的转录由相同的启动子（P_R和P_L）控制。这一转换既不涉及新的σ因子也不是RNA聚合酶识别新的启动子开始新的转录，与其他噬菌体不同，它是由相同启动子控制的转录过程的延伸。

晚早期基因对于裂解周期的继续以及溶源态的建立均很重要。*O*基因和*P*基因编码噬菌体DNA复制所必需的蛋白质，而复制是裂解周期的关键。*Q*基因的产物（Q是另一个抗终止子）可使晚期基因转录。

晚期基因都是沿右向转录的，但不是从P_R开始的。晚期启动子$P_{R'}$恰好位于*Q*的下游。如果Q不干预转录终止，那么从$P_{R'}$开始的转录仅进行194bp就终止了。*N*基因的产物不能代替Q，它对*cro*和*N*基因后的抗终止作用是特异性的。晚期基因编码噬菌体头部和尾部蛋白以及裂解宿主细胞的蛋白质，将子代噬菌体释放出来。

小结 在λ噬菌体裂解周期中，早期/晚早期/晚期基因转录的转换是受抗终止蛋白控制的。*cro*是两个早期基因之一，其编码产物是*cI*基因的阻遏物，从而使裂解周期继续下去。另外一个基因*N*编码抗终止蛋白N，使转录越过*N*和*cro*基因的终止子进入晚早期基因转录。晚早期基因之一的*Q*基因编码另外一个抗终止蛋白（Q），使晚期基因从晚期启动子$P_{R'}$处开始转录，不至于提前终止。

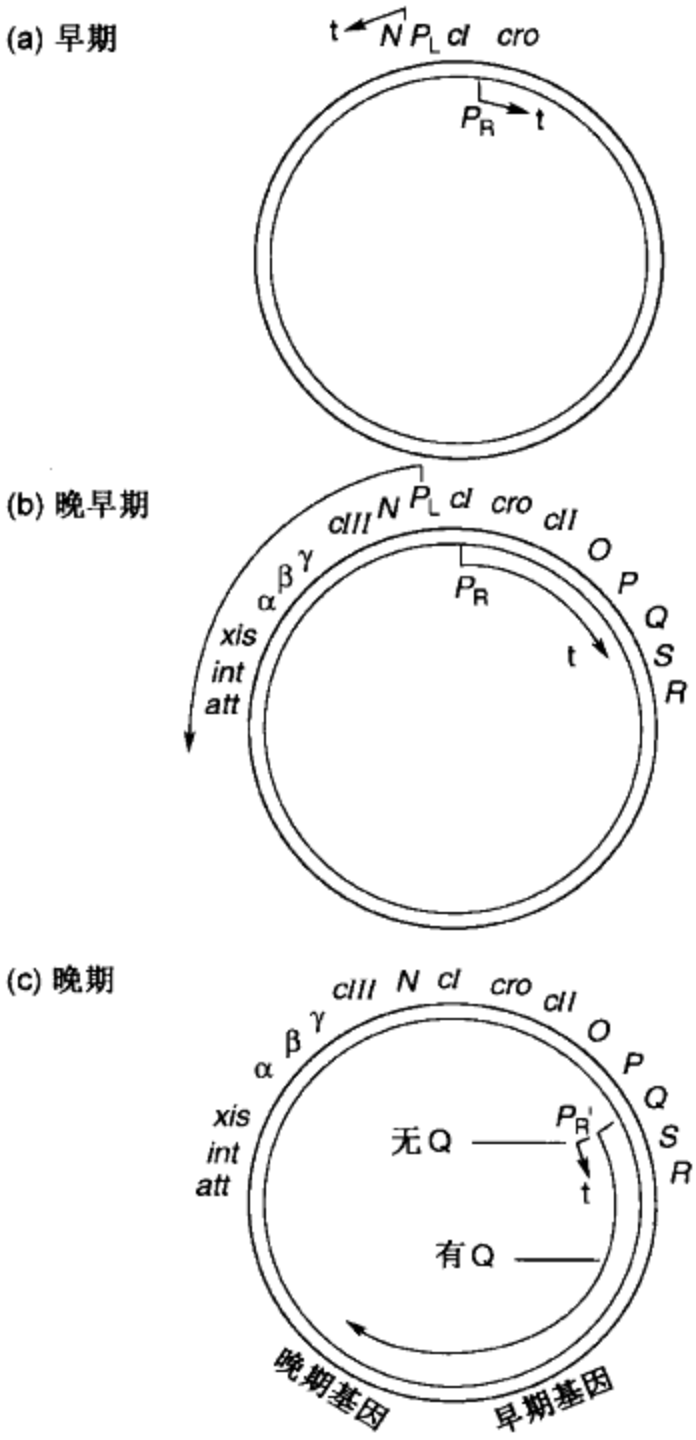


图 8.12　λ噬菌体裂解性感染后转录的时序控制。（a）早期基因转录（红色）始于阻遏物基因 *cI* 两侧的右向和左向启动子（P_R 和 P_L），停止于 *N* 和 *cro* 之后的 rho-依赖型终止子（t）处；（b）晚早期转录（蓝色）在相同的启动子（P_R 和 P_L）处开始，通过抗终止子 N 越过终止子；（c）晚期转录（绿色）在一个新的启动子 P_R' 处开始，如果没有 *Q* 基因的产物 Q（另一个抗终止因子）它就会在终止子 t 处停止。此处 O 和 P 是晚早期基因的编码蛋白，不是操纵子和启动子。

抗终止　N 和 Q 如何行使其抗终止功能呢？它们可能采取两种不同的机制。先看 N 介导的抗终止，图 8.13 显示了这一过程。图 8.13a 中给出 *N* 基因周围的遗传位点。在其右面是左向启动子 P_L 和操纵子 O_L，这里是向左转录的起始处。*N* 基因的下游（左侧）有个转录终止子，在缺乏 *N* 基因产物（N）时，向左的转录会终止。图 8.13b 显示缺乏 N 时所发生的反应。RNA 聚合酶（粉色）在 P_L 处开始转录 *N* 基因，遇到终止子后即从 DNA 上解离，释放 *N* mRNA，N 蛋白合成，图 8.13c 显示接下来所发生的反应。N 蛋白（紫色）与 **N 蛋白利用位点** ***nut* site**（N utilization site，绿色）结合并与已结合于 RNA 聚合酶的宿主蛋白复合体（黄色）相互作用。由此在某种程度上改变了聚合酶活性，使其变得“盲目”而忽略终止子，因而转录持续到晚早期基因。同样，从 P_R 开始向右的转录也采取同样的机制，因为在 *cro* 的右端也有一个位点使聚合酶忽略终止子而转录晚早期基因。

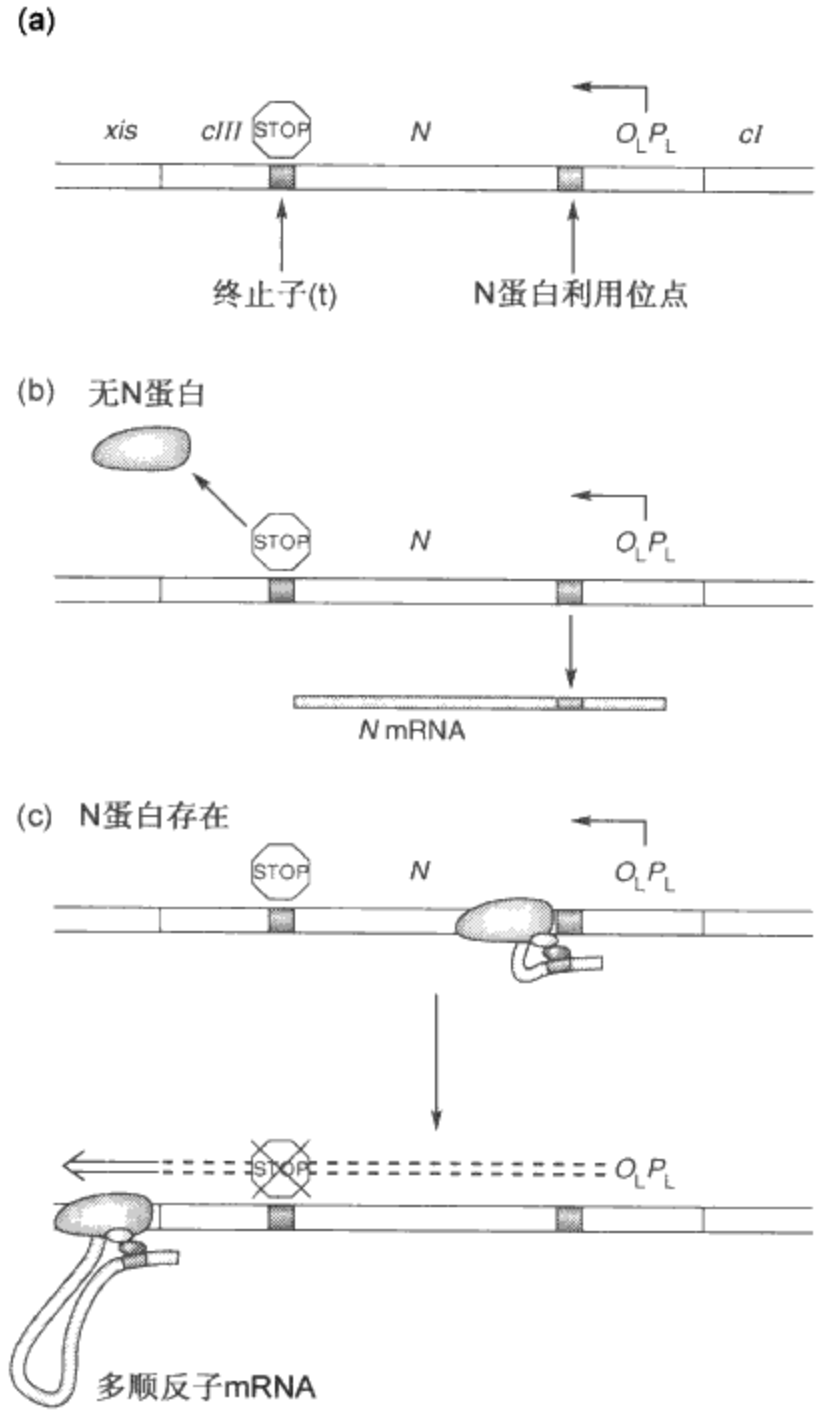


图 8.13　抗终止子 N 对左向转录的影响。（a）λ噬菌体基因组的 *N* 区。*N* 周围的基因分别是左向启动子 P_L 和操纵子 O_L、终止子（红色）、*nut* 位点（绿色）。（b）没有 N 情况下的转录。RNA 聚合酶（粉色）在 P_L 处开始向左转录，在位于 *N* 基因后的终止子处停止。*N* mRNA 是唯一转录产物。（c）N 存在时的转录。N（紫色）与转录本的 *nut* 位点结合并且与 NusA（黄色）结合，而NusA 又与其他蛋白（未显示）一起与 RNA 聚合酶结合成复合体。这一复合体使聚合酶可以通读终止子，继而转录晚早期基因。

怎样得知宿主蛋白也参与抗终止呢？遗传学研究表明 4 个宿主基因的突变会影响抗终止。这些基因编码蛋白质 **NusA、NusB、NusG** 以及核糖体 S10 蛋白。这些宿主蛋白的相互作用会导致宿主细胞死亡，这听起来似乎有些难以置信，其实这只是病毒为了自身利益而控制寄主细胞过程的诸多实例之一。在本例中，S10 蛋白显然与合成蛋白质有关，Nus 蛋白使 7 个 *rrn* 操纵子抗终止，rrn 操纵子编码核糖体 RNA 和 tRNA。实际上，我们将会看到 NusA 促进了终止作用。

体外实验表明 N 和 NusA 这两个蛋白质在终止子距 *nut* 位点足够近时可产生抗终止作用。图 8.14a 显示在这种小范围的抗终止过程中所涉及的蛋白复合体，同时也说明了 N 蛋白自身并不结合在 RNA 聚合酶上，而是与 NusA 结合，然后 NusA 再与聚合酶结合。图中还介绍了 *nut* 位点的两个部分：A 盒（box A）与 B 盒（box B）。A 盒在 *nut* 位点高度保守，而 B 盒则不然，不同的 *nut* 位点间存在差异。B 盒转录本中包含一个反向重复序列，可形成茎环结构，如图 8.14 所示。

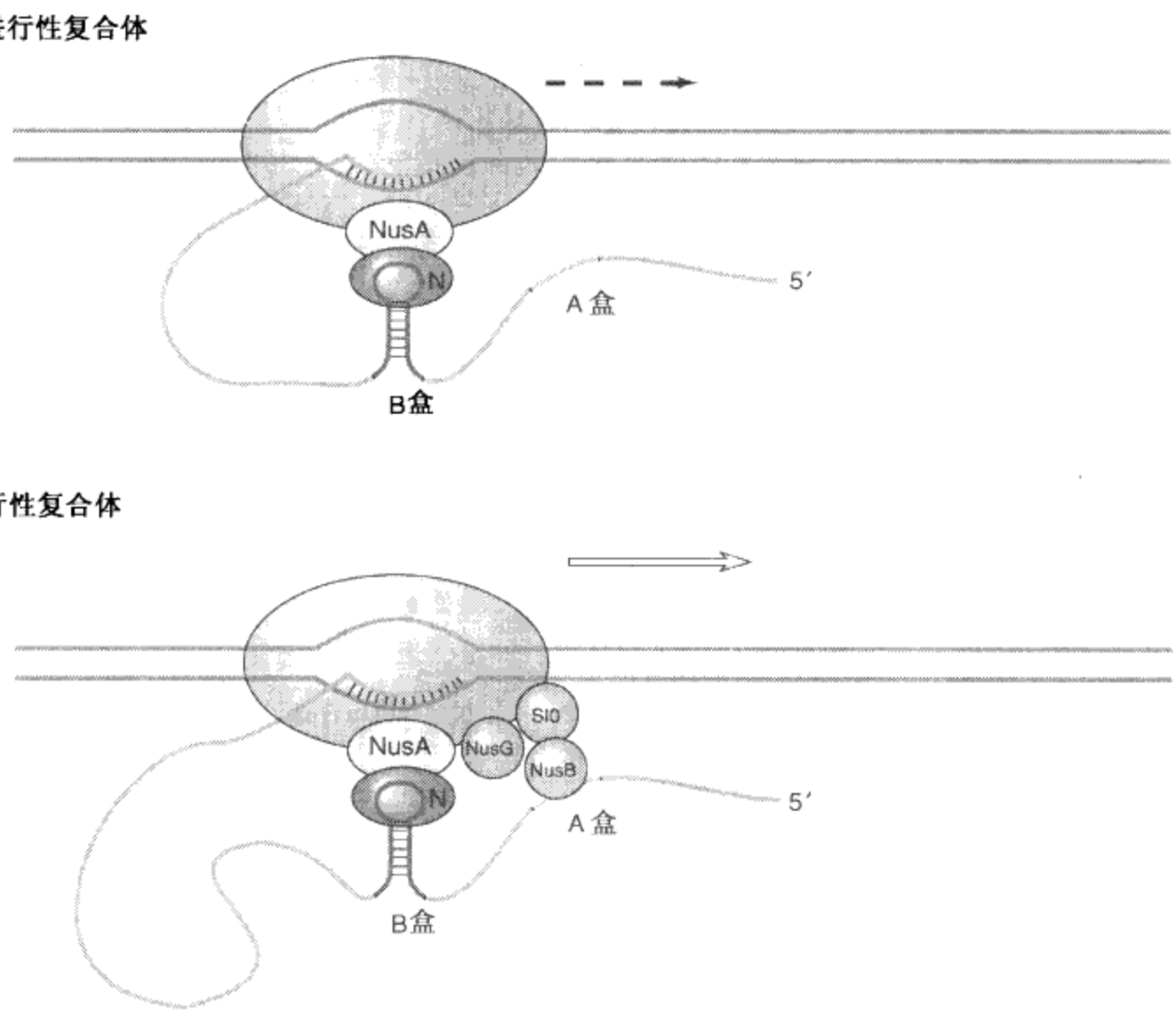


图 8.14　参与 N 介导抗终止的蛋白质复合体。（a）弱的非进行性复合体。NusA 结合到聚合酶上，N 结合到 NusA 和转录本 *nut* 位点的 B 盒，在延伸中的 RNA 上产生一个环。该复合体相对较弱，只在 *nut* 位点附近引起抗终止（虚线箭头）。这种情况只在体外存在。（b）强的进行性复合体。如（a）所示，NusA 将 N 和 B 盒固定到聚合酶上，S10 结合到聚合酶上，NusB 结合到转录本 *nut* 位点的 A 盒上，使聚合酶与转录本的结合更为紧密，复合体更加稳定。NusG 对复合体的加强也有作用。这个复合体是进行性的，在体内可以引起下游数千碱基对的抗终止（空心箭头）。

体内抗终止并不像上述示意过程，因为抗终止发生在相对于 *nut* 位点下游至少数百个碱基对的终止子上。这种抗终止是**进行性的**（processive），因为聚合酶在 DNA 链上移动相当长的距离时，抗终止因子始终与其结合。这种进行性抗终止不仅需要 N 和 NusA，还需要其他 3 种宿主蛋白 NusB、NusG 和 S10。推测这些蛋白质帮助稳定抗终止复合体使其稳定到达终止子。图 8.14b 显示包含所有这 5 个抗终止蛋白的稳定复合体。

对图 8.14 中所述抗终止复合体而言，最预想不到的特征是它不与 *nut* 位点本身相互作用，而是与 *nut* 位点的转录物相互作用。何以得知？其中一个证据是 N 蛋白上识别 *nut* 位点的重要区域富含精氨酸，这与 RNA-结合结构域相似。Asis Das 提供了更直接的证据，他利用凝胶阻滞实验证明 N 与含有 B 盒的 RNA 片段结合。此外，当 N 和 NusA 都与复

合体结合时，可部分地保护 B 盒免受 RNA 酶降解，但不保护 A 盒。只有当所有 5 个蛋白质都结合到复合体时才能保护 A 盒免受 RNA 酶降解。这些结果与图 8.14 的模型一致。

那么如何知道存在图 8.14 中所示的 RNA 环呢？对此还不能肯定，但很容易想到从 N 蛋白结合到 *nut* 位点的转录本开始，直到聚合酶达到终止子为止，如果 N 蛋白始终保持与聚合酶和 RNA 的结合，就需要形成如图所示的 RNA 环。为了验证这个假设，Jack Greenblatt 小组分离到一些突变体，其编码 RNA 聚合酶 β 亚基的基因发生了突变，从而影响 N 蛋白介导的抗终止作用。这些突变体在体外转录中不能保护 *nut* 位点的转录本，表明在转录过程中 RNA 聚合酶、N 蛋白和 *nut* 位点的转录本之间存在着结合。于是很容易再次想象到位于 *nut* 位点的转录本和聚合酶之间的 RNA 会形成一个环。

那么 N 蛋白如何阻止终止呢？有一个假设认为 N 蛋白限制了对终止很重要的 RNA 聚合酶在终止子位点的暂停。但是 Ivan Gusarov 和 Evgeny Nudler 在 2001 年认为 N 蛋白并不完全依靠限制聚合酶的暂停来产生对终止作用的显著影响。他们证明，N 蛋白与注定会形成发夹结构的上游 RNA 结合，从而减缓发夹的形成。没有发夹，终止就不会发生。这里可回顾一下色氨酸操纵子转录衰减作用的机制，其中涉及核糖体停止在衰减子发夹结构的上游部分（第 7 章）。

图 8.15 展示了 Gusarov 和 Nudler 的模型 (2001)（其中一部分在第 6 章讨论过），同时也展示了 NusA 在终止中的作用。当**延伸复合物 EC**（elongation complex）在合成 U 串时，加上第 7 个核苷酸后便暂停下来，这就使准发夹结构的上游部分处于跟 RNA **聚合酶上游结合位点 UBS**（upstream binding site）结合的位置。暂停只持续约 2s，发夹必须在这一时间内形成，否则聚合酶就会移过而不发生终止作用。如果发夹形成，则延伸复合物便被套住，终止发生。NusA 的作用是弱化 UBS 与准发夹上游部分的结合，帮助发夹在暂停结束前形成，以促使终止。

图 8.15 模型还要求 N 蛋白与准发夹上游部分结合以阻止发夹形成。此外，一旦 N 蛋白与 RNA 结合了，它也跟 NusA 结合。NusA 在这里也与准发夹上游部分结合。由于 N 蛋白和 NusA 同时与 RNA 结合，发夹结构形成得非常慢，使聚合酶顺利通过而不会终止。

有什么证据支持这一模型呢？模型的关键部分是 N 蛋白（以及 N+NusA）与 RNA 发夹上游部分的相互作用。Gusarov 和 Nudler 用蛋白质-RNA 交联技术证明它们之间确实有相互作用。在有或无 N 蛋白和 NusA 条件下，通过一次步移或延伸 [^{32}P] RNA 的一至数个核苷酸的方法（见第 6 章），在 RNA 的+45 位引入 **4-硫尿嘧啶（4-thioU**，sU）。然后继续步移，分别得到延伸至+50、+54、+58、+62、+68和+75 位的 RNA，其 sU 位点相对于 RNA 的 3′端，分别位于−6、−10、−14、−18、−24、−31 位。

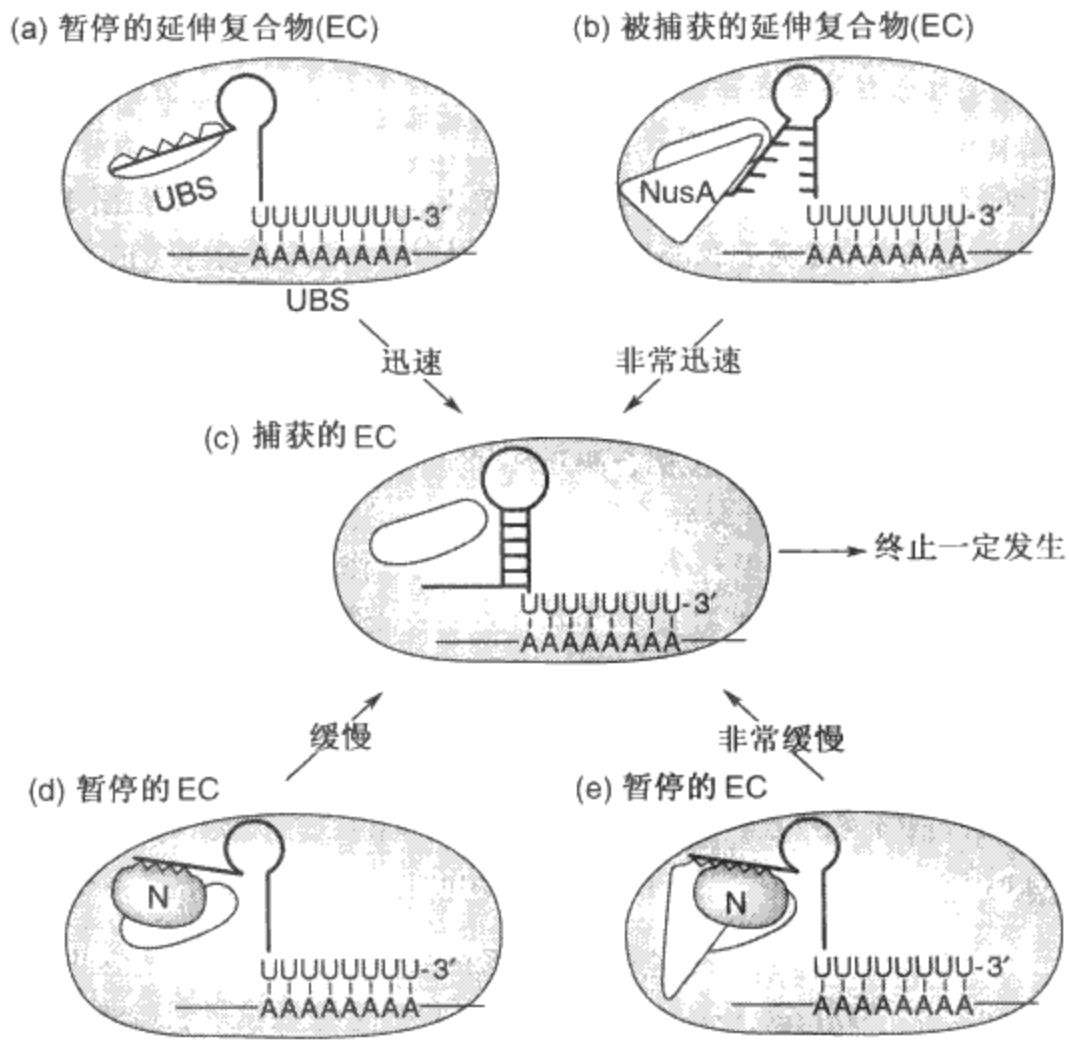


图 8.15 NusA 和 N 蛋白在终止作用中的作用模式。（a）准发夹的上游部分与核心聚合酶的上游结合位点（UBS）结合。然而，蛋白质-RNA 键（锯齿表示）断裂，发夹迅速形成，套住延伸复合物（EC）引发终止（c）。（b）NusA帮助打开 UBS 和准发夹上游部分的结合，加速发夹在暂停结束前形成，促使终止发生。（d）N 蛋白与准发夹上游部分结合（锯齿表示）减缓发夹形成，使终止不易发生。（e）N 蛋白不仅与准发夹上游部分结合，也使 NusA 容易与邻近位置结合，使其同时也与准发夹上游部分结合（锯齿表示），使发夹形成得更慢，终止更不易发生。（*Source*：Adapted from Gusarov I. and E. Nudler. 2001. Control of intrinsic transcription termination by N and NusA: The basic mechanisms. *Cell* 107: 444.）

4-硫尿嘧啶受光激活并在紫外线作用下会与任何紧密相连的蛋白质发生交联（碱基的1Å之内）。所以Gusarov和Nudler用紫外线照射sU位点相对于新生RNA的3′端不同的复合体，然后用SDS-PAGE和放射自显影来检测RNA与N、NusA以及核心聚合酶（α、β以及β′）之间的交联。如果RNA与蛋白质发生交联，则该蛋白质会被放射性标记，放射自显影会显示条带。

图8.16a是实验结果。从新生RNA 3′端算起，sU在−18、−24位时能与N和NusA交联（泳道6、7、2、13），而这里正是发夹上游部分的位置。此外，当N和NusA同时出现时，NusA与RNA结合得更为紧密，而且延伸到−31区。N和NusA在终止发生时结合到发夹的上半部分，表明它们在该位置通过控制发夹的形成控制转录终止。N的一个突变体N^{RRR}，其RNA结合RRR基序发生了突变，但仍可以像野生型一样与RNA发夹结合，表明RNA与发夹结合不需要该基序。

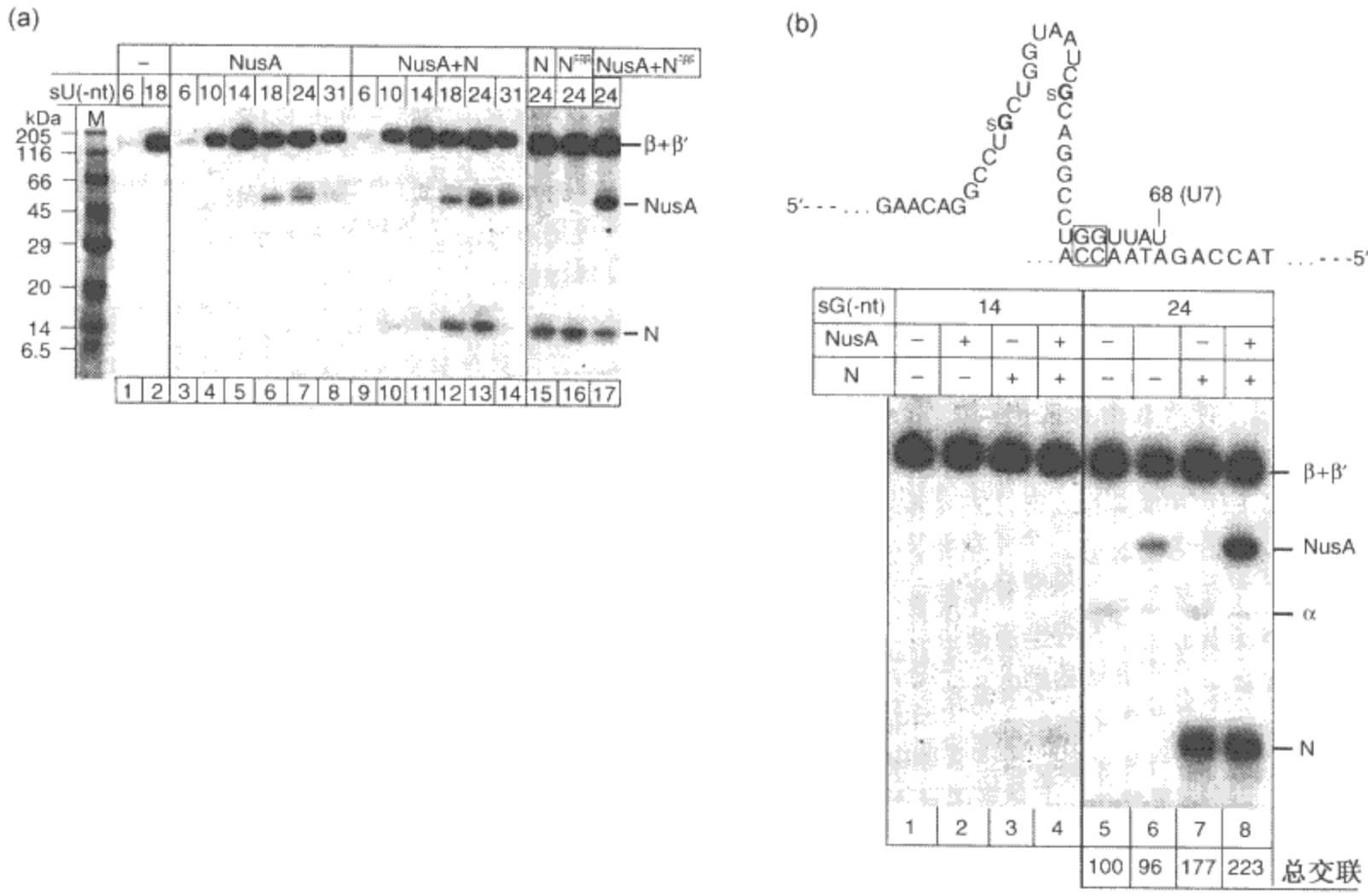


图8.16　在NusA和N有或无的情况下，暂停的EC中蛋白质-RNA结合的证明。（a）N和NusA与终止子上的RNA发夹上游部分的交联。用^{32}P标记新生的含终止子的转录物，通过步移使延伸复合体在+45位参入光激活核苷酸（4-thio-UMP）（见第6章）。然后在NusA、NusA+N、N存在时，分别将复合体延伸至−6、−10、−14、−18、−24、−31位，如顶部所示。然后将复合体暴露在紫外光下使4-thio-UMP与结合到RNA上的蛋白质交联。之后经SDS-PAGE和放射自显影鉴定与RNA共价结合的蛋白质。N、NusA及RNA聚合酶β+β′的位置在右侧指出。泳道M为标准分子质量。N^{RRR}表示N的突变体（RRR RNA结合结构域突变）。N和NusA在−18和−24位都有结合，此处包括发夹的上半部分。（b）N和NusA对聚合酶核心酶与发夹相互作用的影响。实验与（a）相似，但是用6-thio-G（sG）做交联剂，用了一个突变终止子（T7-tR^{mut2}）减缓终止。sG的位置（相对于RNA的3′端−14或−24位）以及N、NusA、N+NusA如顶部所示，蛋白质β+β′、NusA、α以及N的位置在右侧标出。上部的图表示sG在−14或−24位，加框碱基对是终止子上突变的位置。NusA使发夹上半部分（−24位）与核心聚合酶（β+β′和α）的结合减弱，但是N却促进RNA与核心酶结合，而N+NusA则极大地增加了RNA与核心酶的结合（见底部方框中所标的“总交联”）。（*Source*：Reprinted from *Cell* v. 107，Gusarov and Nudler，p. 443 © 2001，with permission from Elsevier Science.）

为了验证N和NusA控制发夹形成这一假设，Gusarov和Nudler还是利用步移法制备了不同的延伸复合体，但是这次将标记的RNA延伸至一个突变的终止子（T7-tR^{mut2}），该终止子在RNA末端位于寡聚尿嘧啶区的两个U被置换成了两个G。这一变化延缓了发夹的形成以便在终止前研究延伸复合

体。他们在－14 或－24 位掺入了另一种光激活核苷酸 6-thioG（sG）。然后在分别有 N、NusA、N＋NusA 情况下进行步移。之后紫外交联，SDS-PAGE 和放射自显影检测 RNA 与蛋白质的结合。

图 8.16b 表明，当 sG 位于－14 位时（发夹的下半部分），RNA 与 N、NusA 都不交联，但是当 sG 位于－24 位时（发夹的上半部分），便有交联发生。因此，N 和 NusA 似乎都与形成发夹上半部分的 RNA 接触。NusA 结合 RNA 后降低了与核心聚合酶（对 α、β、β′亚基，比较泳道 5、6）的结合。相反 N 结合 RNA 后没有降低其与核心聚合酶的结合（比较泳道 5、7）。类似地，当 N 和 NusA 一起结合 RNA 时，不降低发夹上半部分与核心聚合酶的结合（比较泳道 5、8）。这些结果表明，NusA 干扰了聚合酶与发夹上游部分的结合，而 N 则阻止这种干扰并恢复了聚合酶与发夹之间的结合。

Robert Landick 小组进行了相似的交联实验，但是以 5-碘尿嘧啶（5-iodoU）作交联剂，将其置于暂停发夹环的－11 位，引起转录暂停但不终止。他们发现 RNA 与 NusA 的弱结合取代了与 β 亚基的强结合。此外，RNA 发夹环与 RNA 聚合酶 β 亚基的交联发生在 β 亚基的 **flap-tip 螺旋（flap-tip helix）**区。而且，从 flap-tip 螺旋区删除几个氨基酸就破坏了 NusA 所引发的暂停。因此对于 NusA 来说 flap-tip 螺旋是必需的。由于 β 亚基的激活位点与这一 flap 直接相关，他们推测有一个变构机制：暂停的发夹环与 flap-tip 螺旋的相互作用改变了激活位点的构象，使延伸变得更加困难，从而使聚合酶暂停。NusA 可能有利于这一过程。

如果图 8.15 中 Gusarov-Nudler 的模型是正确的，那么在发夹的上下游之间插入一个大 RNA 环就可以干扰 N 和 NusA 的活性。因为 N 和 NusA 与 UBS 的结合正如与发夹下游前面的 RNA 相互结合一样。通常情况下这是发夹的上游部分，而在此处仅为一个大的 $tR2^{loop}$ 环的起始部分。因此，N 和 NusA 不能在准确的位置影响发夹形成，因而对终止的影响甚小。Gusarov 和 Nudler 用 $tR2^{loop}$ 终止子验证了这个假设。正如预期的，N 或 NusA 对终止都没有太大影响。

晚期转录的控制也需利用抗终止机制，但是与 N 抗终止系统完全不同。图 8.17 显示 ***qut* 位点（Q 利用位点，Q utilization site）**与晚期启动子（P_R'）重叠，也与转录起始点下游 16～17bp 处的暂停位点重叠。与 N 系统相反，Q 直接与 *qut* 位点结合而不是与其转录本结合。

缺乏 Q 时，RNA 聚合酶到达转录暂停信号后暂停数分钟。随后越过暂停位点，转录到终止子后

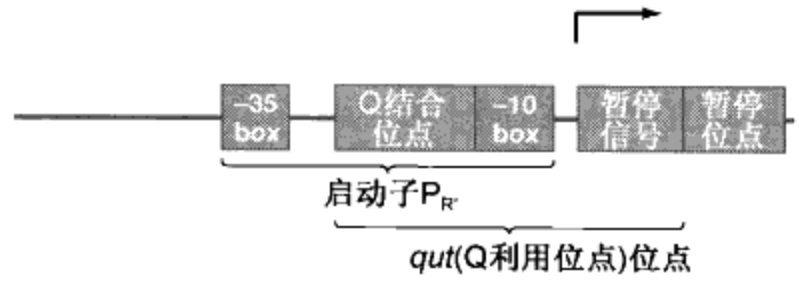


图 8.17　λ 基因组的 $P_{R'}$ 区图谱。启动子贯穿－10 框与－35 框。*qut* 位点与启动子重叠，并包括－10 框上游的 Q 结合位点、转录起始位点下游的暂停信号、＋16 和＋17 位的暂停位点。（*Source*：Adapted from *Nature* 364：403，1993.）

便放弃晚期基因的转录。如果存在蛋白 Q，Q 可识别暂停复合体并与 *qut* 位点结合，然后再结合到聚合酶上，使转录恢复，忽略终止子，继续转录晚期基因。Q 通过抑制紧接在聚合酶之后的 RNA 发夹环的形成来改变聚合酶活性，从而抑制终止子的活性。在 λ 噬菌体晚期控制区内 Q 自身可以引起抗终止，而 NusA 可使这一过程更为有效。

小结　5 种蛋白质（N、NusA、NusB、NusG、S10）在 λ 早期终止子的抗终止作用中相互协作。NusA 和 S10 结合到 RNA 聚合酶上，N 和 NusB 分别结合到正在延伸的 RNA *nut* 位点的 A 盒和 B 盒区。N 和 NusB 分别与 NusA 和 S10 结合，很可能将聚合酶与转录本结合在一起，使聚合酶能通读早期基因的终止子。NusA 通过刺激 RNA 发夹上游部分与核心聚合酶的结合来激活终止作用，从而有利于新生 RNA 发夹结构的形成。发夹不可逆转地改变复合体的结构从而引发终止。N 通过结合到发夹的上游部分阻止发夹形成来干预这一过程。N 甚至可以帮助 NusA 更好地与新生的 RNA 结合从而更有力地抑制发夹形成。没有发夹形成终止就不会发生。λ 噬菌体晚期基因的抗终止作用需要 Q 蛋白与 *qut* 位点结合，使 RNA 聚合酶在晚期启动子的下游停顿。随后，Q 与聚合酶结合使其忽略终止子而继续晚期基因的转录。

溶源态的建立

我们已经提到，晚早期基因不仅对于裂解周期而且对于溶源态的建立都是必需的。晚早期基因通过两条途径帮助溶源态的建立：①某些晚早期基因的产物是噬菌体 DNA 整合到宿主基因组形成溶源态所必需的；②*cII* 和 *cIII* 基因产物使 *cI* 基因转录产生 λ 阻遏物，而 λ 阻遏物是形成溶源态的核心组分。

P_{RM}和 P_{RE}两个启动子控制 *cI* 基因（图 8.18）。P_{RM}代表“阻遏物保持启动子”，确保阻遏物持续供

应以维持溶源态。它有一个特殊性质，要求自身产物阻遏物来激活其自身活性。这一点要进行讨论，但是我们立刻面临一个重要问题，即该启动子不能用来建立溶源态，因为在感染的早期还没有阻遏物来激活它。所幸，启动子 P_{RE} 可解决这个问题。P_{RE} 代表“阻遏物建立启动子”。P_{RE} 位于 *cro* 和 P_R 的右侧，指导向左转录 *cro*，继而是 *cI*。因此 P_{RE} 使 *cI* 基因在阻遏物产生之前表达。

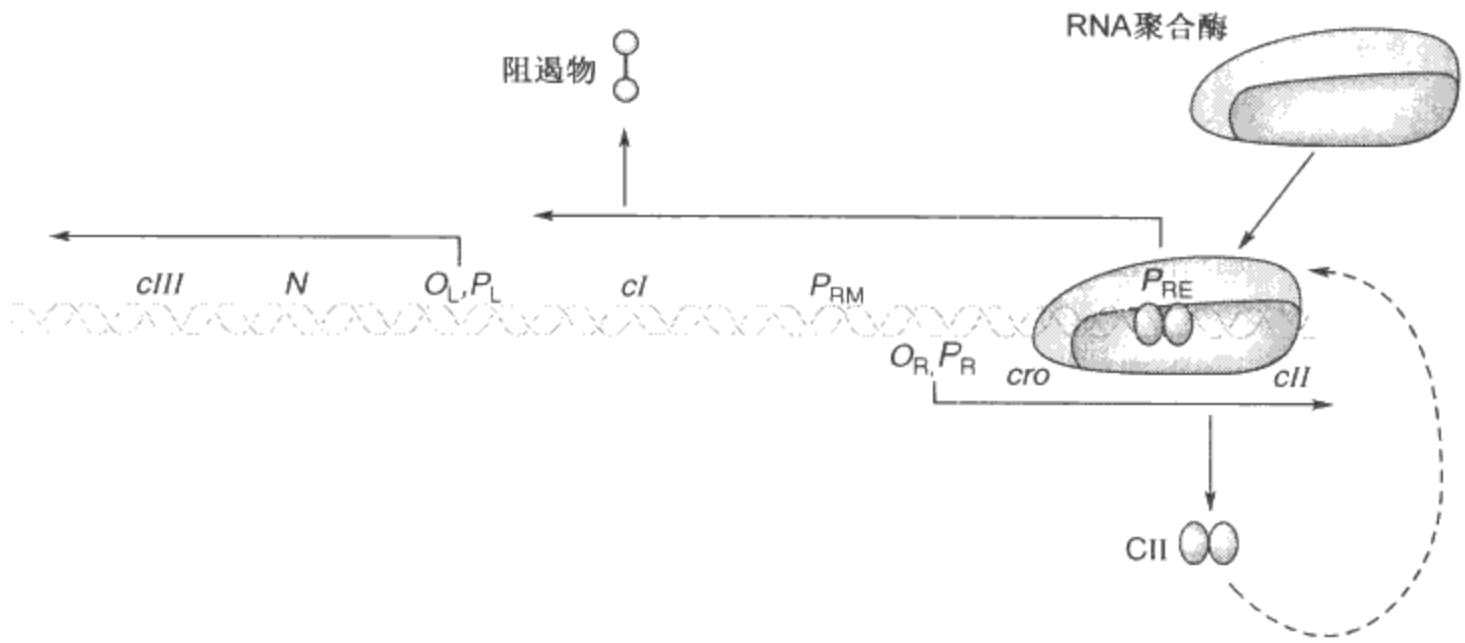


图 8.18　溶源态的建立。从 P_R 转录的晚早期基因产生 *cII* mRNA，继而翻译成 CII 蛋白（紫色）。CII 帮助 RNA 聚合酶（红色和蓝色）结合到 P_{RE} 上，转录 *cI* 基因，产生阻遏物（绿色）。

当然，*cro* 本身的转录方向是从 P_R 开始向右进行的，所以从 P_{RE} 向左转录出 RNA 是 *cro* 的反义（antisense）转录本，也是 *cI* 的“正义”转录本。*cI* 转录被翻译成阻遏物，*cro* 的反义转录本则不能被翻译。这一反义 RNA 有利于溶源态的建立是因为 *cro* 的反义 mRNA 与 *cro* mRNA 相结合，进而干扰其转录。*cro* 抑制溶源态建立，那么阻断其作用可促进溶源态的发生。同样，CII 激活位于 *Q* 内的左向启动子（$P_{anti\text{-}Q}$）的转录，这种“反向”转录产生 *Q* 的反义 RNA，该反义 RNA 封闭了 Q 产生。由于裂解过程的晚期转录需要 Q，而阻碍其合成将有利于另一个途径溶源态的发生。

P_{RE} 对其自身有一些有趣要求。例如，其－10 框和－35 框序列与一般的为 *E. coli* RNA 聚合酶所识别的序列无明显相似性。它无法在体外单独被聚合酶所转录，但 *cII* 基因产物 CII 能帮助 RNA 聚合酶与这种异常启动子结合。

Hiroyuki Shimatake 和 Martin Rosenberg 在体外证明了 CII 的活性。他们用滤膜结合实验显示单独的 RNA 聚合酶或 CII 蛋白都不能与含有 P_{RE} 序列的 DNA 片段结合，因此也就不能使 DNA 结合到硝酸纤维素滤膜上。但 CII 蛋白加上 RNA 聚合酶便可以使 DNA 结合到滤膜上。因此，CII 需激活 RNA 聚合酶使其结合到 P_{RE} 上。而且，这种结合是特异的，CII 激活的聚合酶只能与 P_{RE} 及另一个启动子 P_I 结合。启动子 P_I 控制 *int* 基因，*int* 基因和 CII 是建立溶源态所必需的。*int* 基因还参与 λ 噬菌体 DNA 整合到宿主基因组的过程。

在第 7 章介绍了 CAP-cAMP 通过蛋白质-蛋白质间的相互作用激活 RNA 聚合酶对 *lac* 启动子的结合。Mark Ptashne 小组证明 CII 可能通过相似的途径发挥作用。他们用 DNA 酶印迹实验（见第 5 章）定位 CII 在含 P_{RE} DNA 片段上的结合位点，发现 CII 结合到启动子的－21 和－44 位点间。显然，该区间包含不能识别的启动子－35 框。由此提出一个问题，两个蛋白质（CII 和 RNA 聚合酶）如何在同一时间结合到同一位点上？通过对 CII 的 DMS 印迹实验（见第 5 章）做出如下解释：通过 CII 所结合的这些碱基（如在－26 和－36 位的 G's）位于应与 RNA 聚合酶结合的那些碱基（如－41 位上的 G）的螺旋背面，即这两个蛋白质分别结合在 DNA 螺旋相对的两侧，因此可以合作而不是竞争。

为什么 CII 可单独印迹 λ 噬菌体启动子 DNA，而在滤膜结合实验中却不能与其相结合呢？印迹实验非常灵敏，只要平衡混合物中的蛋白质结合到 DNA 上就能检测到，而且还能保护 DNA。而滤膜结合实验要求较高，一旦蛋白与 DNA 分离，DNA 就流过滤膜不再被重新结合。

cIII 基因的产物 CIII 同样参与溶源态的形成，但其作用是间接的，主要是防止 CII 被细胞内的蛋白酶降解。因此，晚早期基因的产物 CII 和 CIII 通过相互协作激活 P_{RE} 和 P_I 来实现溶源态的建立。

小结 λ噬菌体通过产生大量的阻遏物与早期操纵子结合以阻止早期 RNA 的进一步合成，以此实现溶源态的建立。用于溶源态建立的启动子是 P_{RE}，它位于 P_R 和 *cro* 的右侧。从这个启动子开始向左转录 *cI* 基因。晚早期基因 *cII* 和 *cIII* 的产物也参与溶源态形成过程：CII 直接刺激聚合酶结合到 P_{RE}和 P_I 上，而 CIII 可以减缓 CII 的降解。

溶源期 *cI* 基因的自我调节

一旦 λ 噬菌体的阻遏物出现，就可以结合到 λ 噬菌体操纵子 O_R 和 O_L 上。这种结合具有双管齐下的作用，有利于溶源态的建立。首先，阻遏物关闭早期基因的进一步转录，阻断裂解周期的发展。*cro* 的关闭对于形成溶源态尤为重要，因为 *cro* 的产物（**Cro**）可抑制阻遏物活性。阻遏物的第二个效应是通过激活 P_{RM}来促使自身合成。

图 8.19 是这种自我激活机制的原理，其关键在于 O_R 和 O_L 都被细分为 3 个部分，且每一部分都可结合阻遏物。O_R 更有意思，因为它既控制 *cI* 的向左转录，也控制 *cro* 的向右转录。O_R 区的 3 个结合位点分别为 O_R1、O_R2 和 O_R3，它们与阻遏物的亲和性不同，阻遏物与 O_R1 结合最紧，其次是 O_R2 和 O_R3。O_R1 和 O_R2 与阻遏物的结合是协同性的，意味着一旦阻遏物结合到它所“偏爱的”O_R1 上，将有利于另一个阻遏物与 O_R2 的结合。正常情况下对于 O_R3 没有这种协同性。

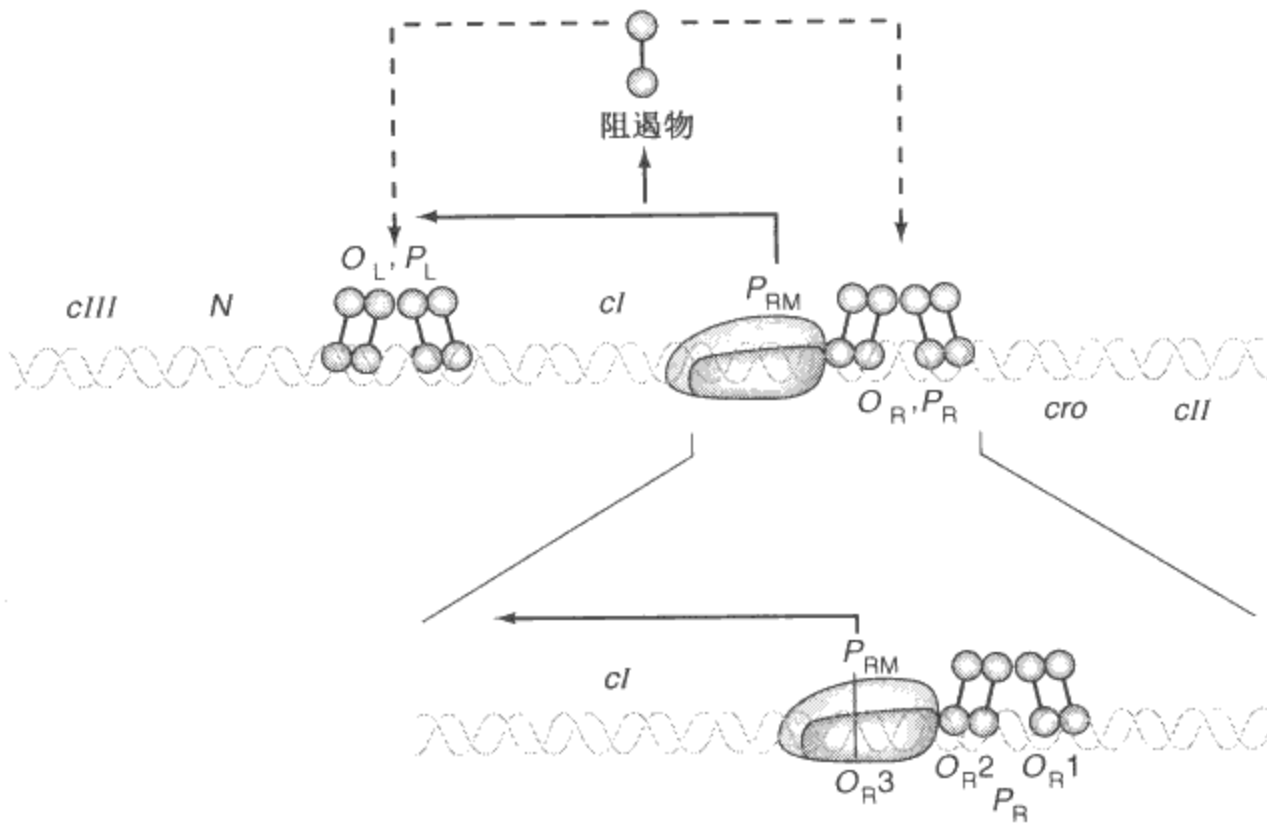


图 8.19 溶源态的维持。（上方）从 P_{RM} 开始的 *cI* mRNA 的转录和翻译产生持续供应的阻遏物，而阻遏物与 O_R 及 O_L 结合，抑制 *cI* 之外任何基因的表达。（下方）控制区详图。阻遏物形成二聚体协同性与 O_R1 和 O_R2 结合。在 O_R2 上的阻遏物与 RNA 聚合酶（红色与蓝色）的蛋白质-蛋白质接触使聚合酶与 P_{RM} 结合转录 *cI*。

阻抑蛋白是由两个相同亚基组成的二聚体，每个亚基在图 8.19 中以一个哑铃形代表。这种形状表明每个亚基在其分子的两端各有一个结构域。这两个结构域拥有不同功能，在 N 端的为 DNA 结合域，C 端的为阻遏物-阻遏物互作位点，该位点使二聚体化及协同作用成为可能。二聚体一旦同时结合到 O_R1 和 O_R2 上，阻遏物便占据 O_R2 十分靠近 RNA 聚合酶在 P_{RM}上的结合位点，使阻遏物与聚合酶紧挨到一起。这种蛋白质-蛋白质间的接触并不发生阻遏效应，而是使 RNA 聚合酶更加有效地结合到这个弱启动子上，这与 CII 促进 RNA 聚合酶结合到 P_{RE}上十分相似。

随着阻遏物结合到 O_R1 和 O_R2 上，便不再从 P_{RE}处转录，因为阻遏物封闭了 *cII* 和 *cIII* 的转录，使 P_{RE} 转录所必需的这些基因产物也被迅速降解。而 cII 和 cIII 的消失却无碍大局，因为溶源态已经形成，仅需少量阻遏物便足以维持溶源性了。只要保持 O_R3 的开启，就可保证所需少量阻遏物的供给，因为 RNA 聚合酶可以从 P_{RM}上自由地转录 *cI* 基因。此外，阻遏物结合到 O_R1 和 O_R2 上，又可通过干扰聚合酶结合到 P_R 上来阻止 *cro* 的转录。

可以想象，阻遏物需要达到较高浓度才能与其最弱结合位点 O_R3 结合。在这种情况下，由于 P_{RM} 被封闭，所有 *cI* 的转录都会停止，而 *cI* 转录的停止使得阻遏物水平下降。阻遏物首先从 O_R3 上解离，从而使 *cI* 转录重新开始。可见这一自我调节机制保证了阻遏物的浓度不至于升得太高。

Ptashne 小组用体外实验证明了以前提出的 3 个假设：①低浓度 λ 阻遏物促进其自身基因的转录；②高浓度阻遏物抑制其自身基因的转录；③阻遏物可抑制 *cro* 的转录。采用一个经过改造的截断转录实验，用低浓度的 UTP 使转录在到达 DNA 模板末端之前暂停。由于转录的暂停导致产生所谓的“打滑”（stut-

ter）转录本。所用模板是一个长为 790bp 的 *Hae*III 限制酶酶切片段，其中含有 *cI* 和 *cro* 基因启动子（图 8.20）。Ptashne 在这篇文章中将 *cro* 基因称为 *tof*。

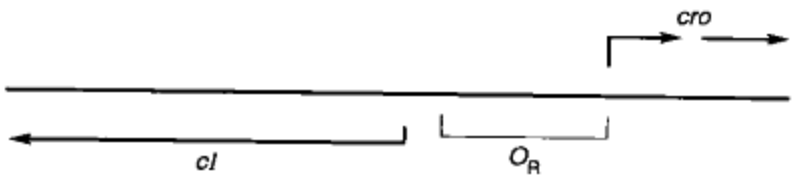


图 8.20　分析从 *cI* 和 *cro* 启动子转录的 DNA 片段图示。红色箭头表示 *cI* 和 *cro* 的体外转录本，其中一些（“打滑”转录本）会提前终止。（*Source*：Adapted from Meyer B. J.，D. G. Kleid，and M. Ptashne. *Proceedings of the National Academy of Sciences* 72：4787，1975.）

他们将 RNA 聚合酶及浓度递增的阻遏物加入模板反应体系，观察 *cI* 基因 300nt 的转录本和另外两个约 110nt 的 *cro* 基因转录本的生成速率。图 8.21 的结果显示，在阻遏物浓度低时，*cro* 转录受到抑制，但 *cI* 的转录受到激发。在阻遏物高浓度时，*cro* 转录彻底停止。而在阻遏物浓度更高时，*cI* 转录也受到抑制。

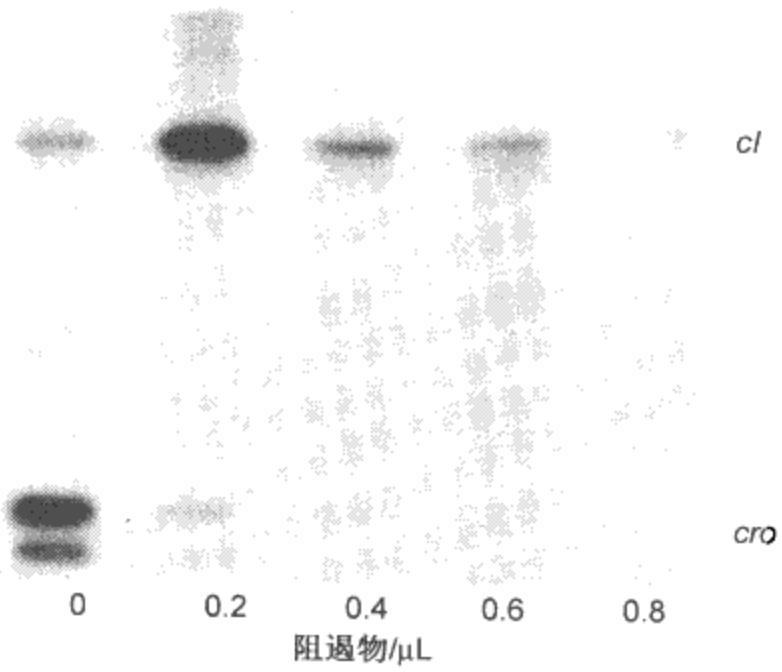


图 8.21　λ 阻遏物对 *cI* 和 *cro* 体外转录影响的分析。Ptashne 小组用图 8.20 所示 DNA 为模板做截断转录（产生“打滑”转录本）。用了不同浓度的阻遏物（见底部）。电泳分离 *cI* 和 *cro* 的打滑转录本（大小如右侧所示）。阻遏物明显抑制 *cro* 转录，但在低浓度下促进 *cI* 转录，在更高浓度下抑制 *cI* 转录。（*Source*：Meyer B. J.，D. G. Kleid，and M. Ptashne. Repressor turns off transcription of its own gene. *Proceedings of the National Academy of Sciences* 72（Dec 1975），f. 5，p. 4788.）

上述体外实验的最大问题是阻遏物浓度远高于溶源菌体内的实际浓度。当 Ptashne 小组在后续实验中采用溶源菌体内阻遏物的生理浓度时发现，它抑制 P_{RM} 对 *cI* 的转录效率只能达到 5%～20%，而要达到 50%的抑制效果，需加入的阻遏物浓度是体内浓度的 15 倍。

基于上述实验结果，很难判断 λ 阻遏物确能抑制自身基因的表达？如果不能，为什么需要 O_R3 和 O_L3？如果它们在正向或负向自我调节中都不起作用，为什么又会存在？Ian Dodd 小组通过检测从 P_{RM} 开始的表达水平发现，仅当 O_L3 存在时，溶源菌体内阻遏物的水平确实可抑制由 P_{RM} 起始的转录（Ptashne 小组在早期实验中没有观察到阻遏物在生理浓度下对 P_{RM} 很强的抑制效果是因为他们构建的载体不含 O_L）。此外，突变研究发现，O_R3 的突变会解除对 P_{RM} 转录的抑制而产生高浓度的阻遏物，并且不能发生溶源态向裂解态的转变（见本章后面部分）。

Dodd 小组以 O_R 和 O_L 间的 DNA 成环理论来解释实验结果，并预测有一个阻遏物八聚体参与其中，如图 8.22 所示。当 DNA 成环时，另一个阻遏物四聚体结合到 O_R3 和 O_L3 上，从而抑制由 P_{RM} 起始的转录。

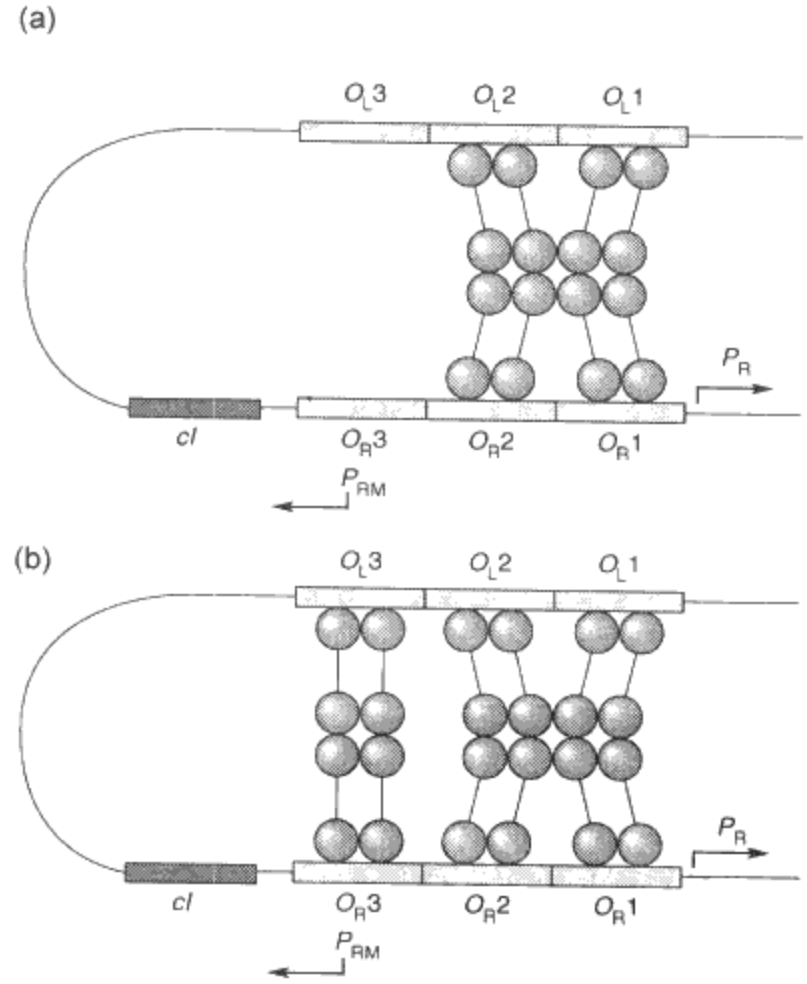


图 8.22　O_L 参与抑制 P_R 和 P_{RM} 的模式。（a）P_R 的抑制。阻遏物八聚体协同性地结合到 O_R1、O_R2 和 O_L1、O_L2 上，使两个启动子间的 DNA 形成环。此时 O_L 对 P_R 的抑制非必需，但为抑制 P_{RM} 奠定基础。（b）P_{RM} 的抑制。由于阻遏物八聚体的形成，阻遏物四聚体协同性地结合到 O_R3 和 O_L3 上，引起对 P_{RM} 的有效抑制，这种抑制作用单靠一个阻遏物二聚体结合 O_R3 是不可能的。

小结 P_{RM}是维持溶源态的启动子。它发挥作用是在由 P_{RE}起始的转录发生之后，该转录使阻遏物迅速合成，形成溶源态。这些阻遏物协同性地结合到 O_R1 和 O_R2 上，但留下 O_R3 敞开。RNA 聚合酶结合到 P_{RM} 上，而 P_{RM} 与 O_R3 重叠，使结合到 O_R2 上的阻遏物与聚合酶接触。这种蛋白质-蛋白质相互作用是启动子有效工作所必需的。高浓度阻遏物抑制从 P_{RM} 开始的转录，而这一过程可能涉及阻遏物二聚体间的相互作用，即结合到 O_R1、O_R2、O_R3 的阻遏物二聚体通过 DNA 环与结合到 O_L1、O_L2 和 O_L3 的阻遏物二聚体相互作用。

RNA 聚合酶/阻遏物相互作用 为什么激发 P_{RM}需要 λ 阻遏物与 RNA 聚合酶的相互作用呢？1994 年 Miriam Susskind 小组的遗传学实验为这一假设提供了强有力的证据。他们获得了一个 λ 噬菌体的突变体，其阻遏物中一个关键的天冬氨酸突变为天冬酰胺。该突变体属于正调控类型（称作 *pc*）。突变的阻遏物可以结合 λ 操纵子，同时抑制从 P_R 和 P_L 开始的转录，但是不能激发从 P_{RM} 开始的转录（即对 *cI* 的正调控不起作用）。Susskind 小组据此做出如下推论：如果阻遏物与聚合酶间的直接作用对高效启动 P_{RM}转录是必需的，那么在 RNA 聚合酶的一个亚基上发生补偿性氨基酸突变应该能恢复与突变阻遏物的相互作用，相应地也能恢复从 P_{RM} 开始的转录。图 8.23 阐述了这种称为**基因间抑制**（intergenic suppression）的机制，因为一个基因的突变可抑制另一基因的突变。

这种基因间抑制突变体的研究，如果对每个部分都分别进行突变研究以了解其活性变化显然极其耗时，而利用筛选（见第 4 章）排除野生型聚合酶基因或排除不相关突变体，只保留想要的突变体会使研究更加有效。Susskind 小组采用了图 8.24 所述的筛选法。他们用了带有两个原噬菌体的 *E. coli* 细胞，其中一个是 λ 原噬菌体含有 *pc* 突变的 *cI* 基因（图中黑色 **X** 所示），该 *cI* 基因的表达受一个弱 *lac* 启动子驱动。另一个是 P22 原噬菌体，含有受 P_{RM} 启动子控制的卡那霉素抗性基因。在有卡那霉素条件下培养细菌，细胞的存活取决于卡那霉素抗性基因的表达。在由细菌提供突变阻遏物和 RNA 聚合酶的情况下，这些细菌由于无法激活从 P_{RM} 开始的转录而无法存活。

接下来，研究者将含有野生型和突变型 *rpoD* 基因的质粒转入 *E. coli*，*rpoD* 基因编码 RNA 聚合酶的 σ 因子。如果 *rpoD* 基因是野生型或无关突变型，则其产物（σ）不会与突变阻遏物作用，因而细菌不能在含有卡那霉素的培养基中生长。如果含有

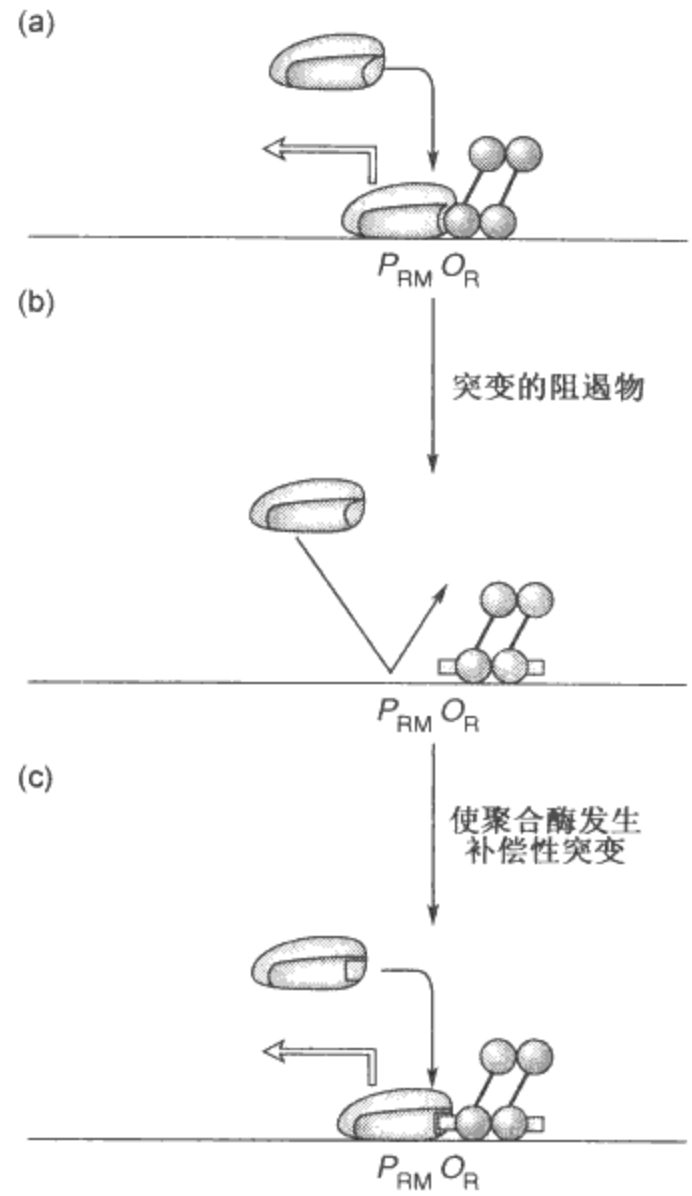


图 8.23 检测 λ 阻遏物与 RNA 聚合酶相互作用的基因间抑制原理。（a）野生型阻遏物与聚合酶紧密结合，激发聚合酶与 P_{RM} 结合并转录；（b）突变的阻遏物基因产生氨基酸发生改变（黄色）的阻遏物，不能与聚合酶结合；（c）突变的聚合酶亚基基因，其中一个氨基酸发生改变（方形凹槽表示），恢复与突变阻遏物的结合，聚合酶与阻遏物现在可以相互作用，由 P_{RM}起始的转录恢复。

阻遏物突变，则其 σ 因子与核心聚合酶结合，形成突变聚合酶可与突变阻遏物相互作用。这种结合可激活从 P_{RM} 开始的转录，细菌因而能在卡那霉素条件下生长。总之，只有 *rpoD* 基因发生阻遏物突变的细菌才可生长，因而很容易找出。

Susskind 小组再次验证了他们突变的 *rpoD* 基因，方法如同筛选时一样，将突变基因转化细菌，只是用 P_{RM}启动子控制 *lacZ* 基因而不是卡那霉素抗性基因。*lacZ* 的产物 β-半乳糖苷酶可测定从 P_{RM} 开始的转录，因而也可分析聚合酶与阻遏物的相互作用。正如所预计的，用突变 *rpoD* 基因转化含突变阻遏物的细菌产生了 120 单位的 β-半乳糖苷酶，这与野生型 *cI* 和 *rpoD* 细菌产生的 β-半乳糖苷酶相当（100 单位）。相反，用野生型 *rpoD* 基因转化突变阻遏物细菌，只产生 18.5 个单位的 β-半乳糖苷酶。对

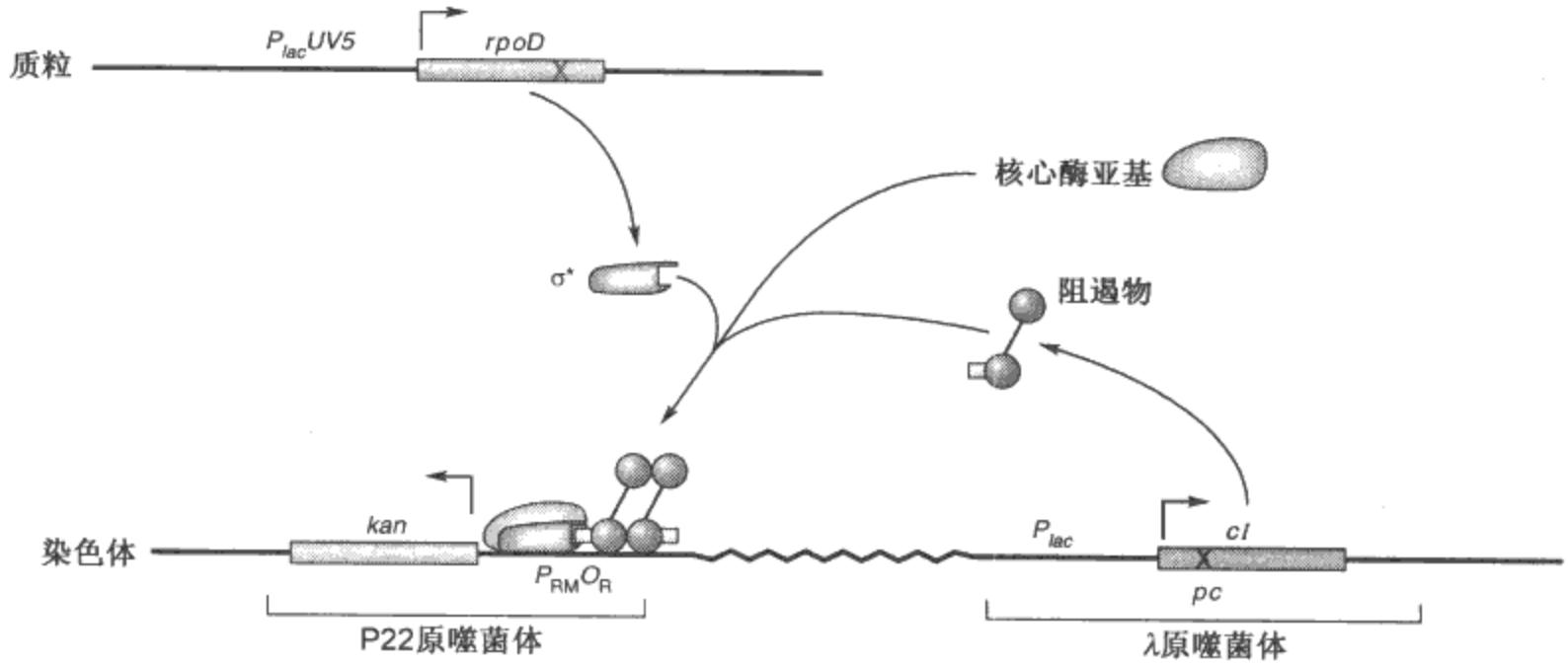


图 8.24　λ*cI pc* 突变基因间抑制物的选择作用。Susskind 小组所用细菌的基因组（图中底部所示的一小部分）包括两个原噬菌体：①P22 原噬菌体带有卡那霉素抗性基因（*kan*），由 λ P_{RM} 启动子及相邻的 λ O_R 启动转录。②λ 原噬菌体含有 λ *cI* 基因（*cI*），由一个弱的 *lac* 启动子启动转录。将带有受 UV5 启动子起始的突变的 *rpoD*（σ因子）基因（*rpoD*）的质粒转化细菌，然后将转化的细菌转入含有卡那霉素的培养基。带有野生型 *rpoD* 或无关突变的 *rpoD* 细菌无法生存，而带有补偿 *cI* 突变（黑色 X）的 *rpoD* 突变（红色 X）体可以生长。这种突变抑制作用可由突变的 σ 因子（蓝色）与突变的阻遏物（绿色）间的相互作用说明，使 *kan* 可以从 P_{RM} 开始转录。

所分离的 8 个阻遏物突变体的 *rpoD* 基因测序发现，在 σ 因子 596 位的精氨酸都突变成了组氨酸，而这里是 σ 因子的第 4 区，它识别－35 框。

这些结果强有力地证实了这一假设：聚合酶与阻遏物的相互作用是激活 P_{RM} 转录的关键。他们还证明这两个蛋白质间激活性的相互作用涉及 σ 因子而不是我们可能会预期的 α 亚基。这为图 8.25 所示通过 σ 因子的启动子激活提供了例证。受这种激活作用的启动子含有一个弱的几乎不能为 σ 因子所识别的－35 框。与－35 框重叠的激活因子位点（activator site）使激活因子（在这里是 λ 阻遏物）处于与 σ 因子的 4 区相互作用的位置，实际上替代了弱识别性的－35 框。

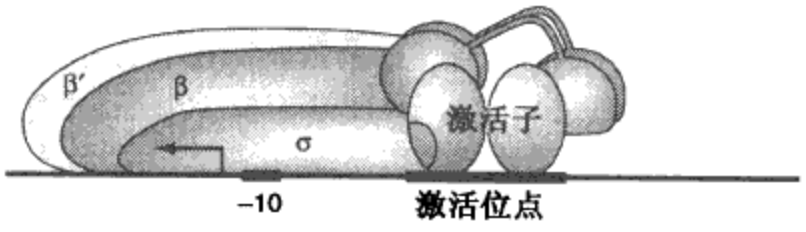


图 8.25　通过 σ 因子接触的激活作用。激活因子（即 λ 阻遏物）结合到激活因子位点上，该位点与启动子的弱－35 框重叠。本来 4 区与－35 框的结合很弱，而激活因子与 σ 因子的 4 区发生相互作用使聚合酶能与弱启动子紧密结合，成功转录邻近的基因。（*Source*：Adapted from Busby S. and R. H. Ebright, Promoter structure, promoter recognition, and transcription activation in prokaryotes. *Cell* 79：743，1994.）

我们在第 7 章得知 CAP-cAMP 通过促进闭合启动子复合体的形成将 RNA 聚合酶召集到 *lac* 启动子上。但是 Diane Hawley 和 William McClue 证明 λ 阻遏物并不影响在 P_{RM} 处的这一步骤，而是促进召集过程的第二步，即帮助将 P_{RM} 处的闭合启动子复合体转换成开放启动子复合体。

小结　对基因间抑制突变的研究显示，阻遏物与 RNA 聚合酶间的关键相互作用涉及聚合酶 σ 亚基的 4 区。这条多肽结合在靠近弱－35 框不远处的 P_{RM} 上，使 σ 的 4 区与结合到 O_R2 上的阻遏物靠近。因此阻遏物可与 σ 因子相互作用，帮助 RNA 聚合酶结合到弱启动子上。通过这种方式，O_R2 充当激活因子位点，而 λ 阻遏物是从 P_{RM} 转录的激活因子，这样就促使闭合启动子复合体转换成开放启动子复合体。

λ 噬菌体感染后寄主命运的决定：裂解或溶源

被 λ 噬菌体感染的细菌是进入裂解态还是溶源态由什么决定？两者间的平衡是精细的，我们不可能预测一个特定细菌的命运。一组受 λ 噬菌体感染的细菌实验证实了这一说法。将少量噬菌体颗粒喷洒在平皿的菌落上，细菌被感染。如果发生裂解性感染，子代噬菌体会感染邻近的细胞，数小时后我们会看到由于细菌裂解而在平皿上留下的圆洞，称为**噬菌斑**（plaque）。如果感染是 100% 裂解性的，

噬菌斑会很清晰，因为所有的宿主细胞都死了。但λ噬菌斑通常都不清晰，甚至比较混浊，表明有溶源菌存在。说明即使在噬菌斑的包围下，被感染的细胞仍然有些进入裂解周期而有些进入溶源态。

让我们稍稍离题来讨论这样一个问题：为什么溶源菌不被噬菌斑上众多的噬菌体裂解性地感染呢？答案是如果一个新的噬菌体 DNA 进入溶源菌，溶源菌中大量的阻遏物就会结合到新的噬菌体 DNA 上阻止其表达。因此可以说，溶源菌对于新噬菌体的**超感染**（superinfection）具有**免疫力**（immune），因为它拥有与原噬菌体相同的控制区或**免疫区**（immunity region）。

现在让我们再回到主题上来，噬菌斑里的有些细菌被裂解性地感染了，而有些却进入了溶源态。噬菌斑上所有细菌的遗传特性是一致的，噬菌体也是如此，显然命运的选择不是遗传性的而是取决于两个基因（*cI* 和 *cro*）产物的合成速度。这种竞争发生在每个被感染的细胞中，胜者决定细菌被感染后的命运。如果 *cI* 先表达则建立溶源态，如果 *cro* 胜出，则被感染的细菌会裂解。我们已经知道这一竞争的基础：如果 *cI* 基因产生足够多的阻遏物，那么这些蛋白质会结合到 O_R 和 O_L 上，阻止早期基因的进一步转录，当然也阻止了能产生子代噬菌体并导致细菌裂解的晚期基因的表达。但是如果产生了足够的 Cro 蛋白，Cro 会阻止 *cI* 基因的转录从而阻止溶源态的建立（图 8.26）。

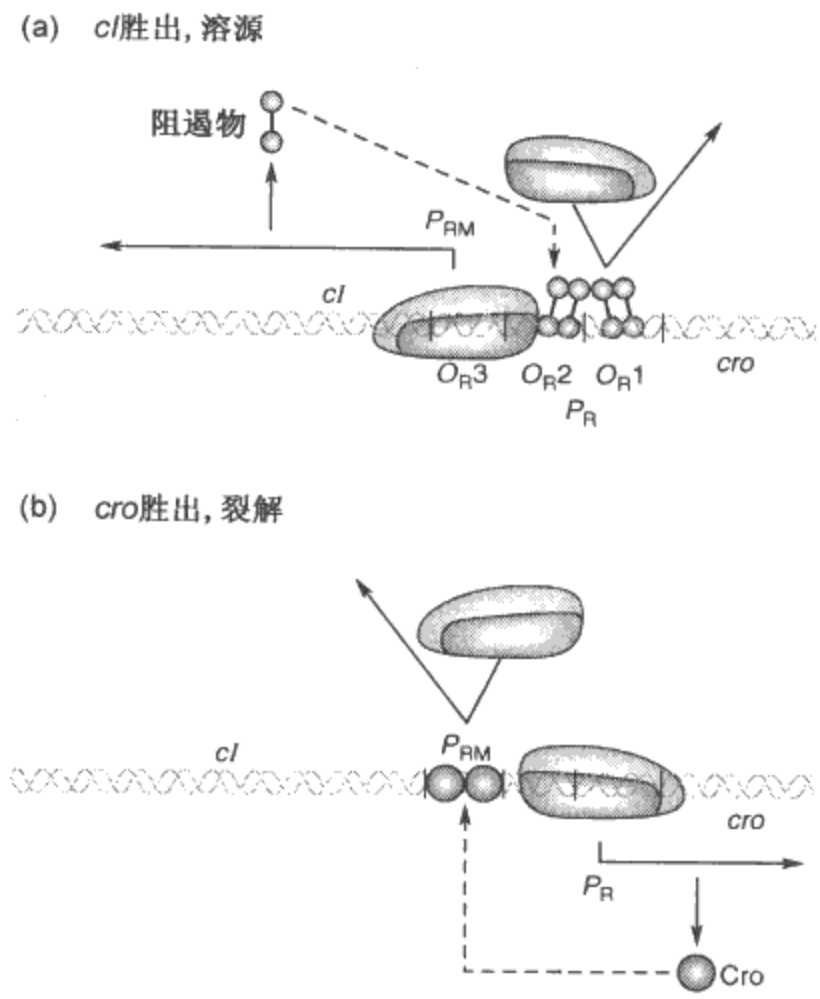


图 8.26 *cI* 与 *cro* 间的战争。（a）*cI* 胜出。从 P_{RM}（和 P_{RE}）转录 *cI* 产生足够多的阻遏物（绿色）阻断聚合酶（红色和蓝色）与 P_R 结合，阻断 *cro* 转录，建立溶源态。（b）*cro* 胜出。从 P_R 开始转录 *cro* 并产生足够多的 Cro（紫色）阻断聚合酶与 P_{RM} 结合，阻断 *cI* 转录，进入裂解态。

Cro 蛋白阻止 *cI* 转录的关键是其与λ操纵子的亲和力。同阻遏物一样，Cro 结合到 O_R 和 O_L 上，但与操纵子 3 个区的结合顺序与阻遏物完全相反。不像阻遏物那样按 1、2、3 顺序，Cro 先与 O_R3 结合。一旦结合后，从 P_{RM} 开始的 *cI* 转录就停止了，因为 O_R3 与 P_{RM} 有重叠。换言之，Cro 充当了阻遏物。此外，当 Cro 的浓度水平达到能与所有左向和右向操纵子都结合的程度，它就阻断了所有从 P_R 和 P_L 开始的早期基因的转录，也包括 *cII* 和 *cIII* 基因。没有这些基因的产物，P_{RE} 就无法发挥功能。结果，所有阻遏物的合成停止，形成裂解性感染。Cro 对早期基因转录的关闭是裂解生长所必需的。在感染晚期，晚早期蛋白质的持续形成可阻断裂解周期。

但是由什么来决定是 *cI* 还是 *cro* 在竞争中胜出呢？显然这不是一个抛硬币问题。实际上最重要的因素还是 *cII* 基因产物（CII）的浓度。细菌内 CII 的浓度越高，越倾向于形成溶源态。这与我们学过的 CII 激活 P_{RE}，帮助开启溶源程序的作用是一致的。我们还知道由 CII 激活的 P_{RE} 通过产生 *cro* 的反义 RNA 抑制正义 *cro* RNA 的翻译来阻止进入裂解程序。

那么是什么控制着 CII 的浓度呢？我们知道CIII保护 CII 免受细胞内蛋白酶的降解，但是高浓度蛋白酶可以不受 CIII 的抑制而降解 CII，也保证了感染后细胞的裂解。例如，在富营养培养基中这种蛋白酶保持较高浓度，而在饥饿条件下，蛋白酶的浓度会降低。因此，饥饿条件下细菌倾向于溶源，而在富营养培养基中细菌倾向于裂解。这对噬菌体是有益的，因为裂解途径需要相当多的能量来生产噬菌体 DNA、RNA 以及蛋白质，而在饥饿情况下则不可能有这么多能量供应。比较而言，溶源态很经济，溶源态建立后仅需合成少许阻遏物。

小结 特定细菌被λ噬菌体感染后是裂解还是溶源取决于 *cI* 和 *cro* 基因产物生成速率的竞争。*cI* 基因编码阻遏物，封闭 O_R1、O_R2、O_L1 及 O_L2，同时关闭所有早期基因包括 *cro* 基因的转录，结果导致溶源。*cro* 基因编码 Cro 蛋白，封闭 O_R3 和 O_L3 从而关闭 *cI* 基因的转录，导致细菌裂解。哪个基因产物首先达到足够高的浓度来阻止其竞争者的合成，它就在竞争中胜出从而决定细菌的命运。竞争的胜出由 CII 的浓度决定，而 CII 的浓度由细胞内蛋白酶的浓度决定，蛋白酶的浓度又相应地由环境因素如培养基的营养度决定。

溶源菌的诱导

我们讨论过化学诱变剂或射线可诱导产生溶源菌。其诱导的机制如下：*E. coli* 细胞应答诱变剂或射线所造成的 DNA 损伤诱导一组基因的表达。这些基因总的功能称为 **SOS 响应**（SOS response），其中最重要的基因是 ***recA***，其产物（**RecA**）参与 DNA 损伤的重组修复（见第 20 章）。这一功能部分地解释了 SOS 响应机制。但是环境胁迫还诱导RecA 蛋白产生一个新的辅蛋白酶（coprotease）活性，激活一些在 λ 阻遏物中未表达的蛋白酶或蛋白质的切割活性。蛋白酶将阻遏物切开，并从操纵子释放，（图 8.27）。一旦切割发生，从 P_R 和 P_L 的转录就开始了。第一个转录的是 *cro*，其产物关闭阻遏物基因 *cI* 的进一步转录，溶源态被瓦解，裂解性噬菌体复制开始。

显然，如果对 λ 噬菌体本身没有益处，不会进化出阻遏物切开自身来响应 RecA 机制。其益处在于 SOS 响应溶源菌遭受 DNA 损伤的信号，帮助原噬菌体进入裂解周期而脱离不利环境，就好比老鼠逃离即将沉没的船一样。

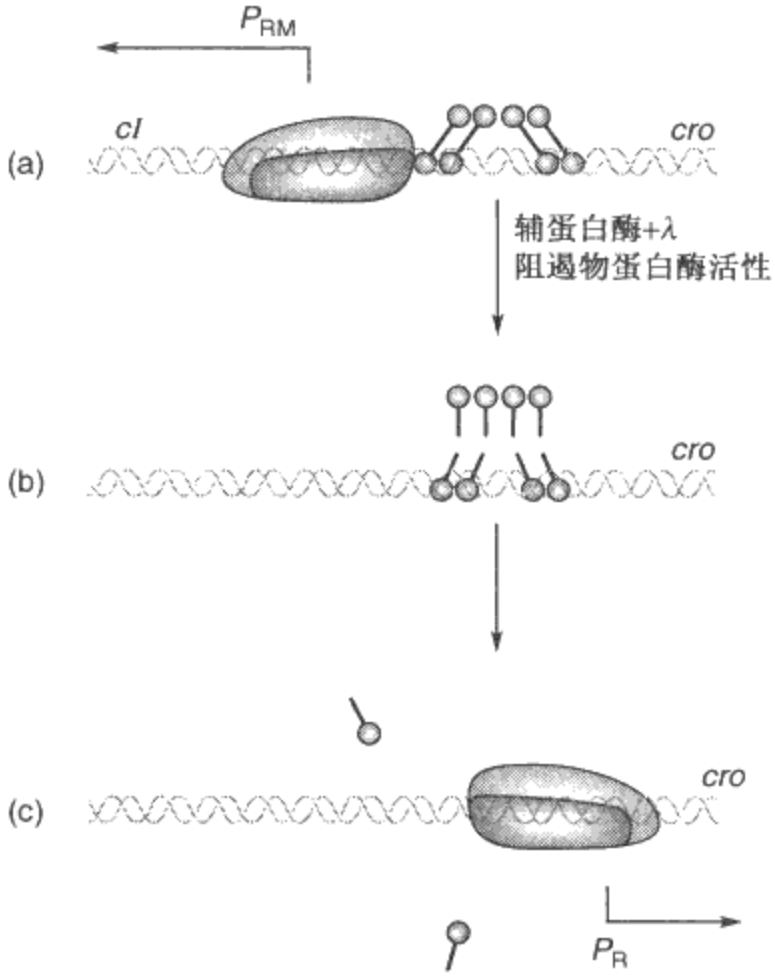


图 8.27　λ 原噬菌体的诱导。（a）溶源态。阻遏物（绿色）结合到 O_R（和 O_L）上，*cI* 从 P_{RM} 启动子开始转录。（b）RecA 辅蛋白酶（由紫外线或其他诱变剂激活）激活阻遏物中的蛋白酶活性，切割自身。（c）被切开的阻遏物从操纵子上脱落，使聚合酶（红和蓝）与 P_R 结合，转录 *cro*，溶源态瓦解。

小结　当溶源菌遭受 DNA 损伤后，会诱导 SOS 响应。这一反应最初是 RecA 蛋白出现辅蛋白酶活性，使阻遏物将自身一切两半，从 λ 操纵子释放，最终诱导裂解周期。子代 λ 噬菌体以这种方式逃避宿主内的致死性损伤。

总结

细菌在转录模式上经历许多重要转变（如在噬菌体感染或孢子发生时），许多机制参与影响这种转变。例如，在被感染的枯草芽孢杆菌细胞中，噬菌体 SPO1 基因的转录是时序性的，即早期基因先转录，接着是中期基因，最后是晚期基因。这种转换由一组噬菌体编码的 σ 因子与宿主核心 RNA 聚合酶共同作用，特异性地从早期到中期再到晚期改变其特异性。宿主的 σ 因子对噬菌体的早期基因具有特异性，噬菌体的 gp28 将特异性转换到中期基因，而噬菌体的 gp33 和 gp34 将特异性转换到晚期基因。

当枯草芽孢杆菌形成孢子时，一整套新的产孢特异基因被启动，此时许多（但不是所有）营养生长基因被关闭。这种转换主要发生在转录水平，这一过程的完成由若干新的 σ 因子取代了营养生长 σ 因子，从而使产孢特异性基因的转录取代了营养生长基因的转录。

有些原核基因在两个不同 σ 因子都有活性情况下必须转录，因此这些基因有两个不同的启动子，分别由两个 σ 因子各识别一个。这样，无论细胞有哪种 σ 因子都可以保证基因的表达，并能在不同条件下进行不同的调控。在 *E. coli* 中，对热激、低氮、饥饿等胁迫的响应都是由 σ 因子转换所决定，分别由 σ^{32}（σ^H）、σ^{54}（σ^N）和 σ^S 取代 σ^{70}（σ^A）来指导 RNA 聚合酶对另一个启动子的选择。

T7 噬菌体没有采用编码新 σ 因子来改变宿主聚合酶对早期基因到晚期基因的特异性这一策略，而是编码一个新的对晚期基因具有绝对特异性的RNA 聚合酶。该聚合酶由一条多肽链构成，是噬菌体最早期基因之一（基因 1）的产物。T7 噬菌体感染后的时序模式比较简单，宿主聚合酶转录最早期基因（I 型），产物之一是噬菌体的聚合酶，然后由噬菌体聚合酶转录晚期基因。

在 λ 噬菌体的裂解周期中，早期/晚早期/晚期基因转录的转换受抗终止机制控制。两个早期基因 *cro* 和 *cI* 中，*cro* 编码对 *cI* 表达的阻抑蛋白 Cro，确保对裂解周期的选择。另一个早期基因 *N* 编码一个抗终止蛋白 N，这个抗终止蛋白可以帮助转录越过 *N* 基因和 *cro* 基因后面的终止子，使转录持续到晚早期基因。晚早期基因中的 *Q* 基因编码抗终止蛋白 Q，使由晚期启动子 $P_{R'}$ 开始转录的晚期基因继续转

录下去而不至于过早终止。

5 种蛋白质（N、NusA、NusB、NusG 和 S10）在 λ 早期的抗终止作用中相互协作。NusA 和 S10 结合到 RNA 聚合酶上，N 和 NusB 分别结合到正在延伸的转录本上 *nut* 位点的 A 盒和 B 盒区。N 和 NusB 分别与 NusA 和 S10 结合，很有可能将聚合酶与转录本结合在一起，使聚合酶能通读早期基因的终止子。NusA 通过刺激 RNA 发夹结构的上游部分与核心聚合酶之间的结合来激活终止作用，从而有利于新生 RNA 发夹结构的形成。发夹结构能够不可逆转地改变复合体结构从而引发终止。N 通过结合到发夹结构的上游部分阻止发夹的形成来干预这一过程。N 甚至可以帮助 NusA 更好地与新生的 RNA 结合从而更有力地抑制发夹结构形成。没有发夹的形成终止就不会发生。λ 噬菌体晚期基因的抗终止作用需要 Q 与 *qut* 位点的结合，使 RNA 聚合酶在晚期启动子的下游停顿。随后，Q 与聚合酶结合使其忽略终止子而继续晚期基因的转录。

λ 噬菌体通过产生大量的阻遏物 CI 与早期操纵子结合以阻止早期 RNA 的进一步合成，来实现溶源态的建立。用于溶源态建立的启动子是 P_{RE}，它位于 P_R 和 *cro* 的右侧。从这个启动子开始向左转录 *cI* 基因。晚早期基因 *cII* 和 *cIII* 的产物也参与溶源态的形成过程：CII 直接刺激聚合酶结合到 P_{RE}和 P_I 上；而 CIII 可以减缓 CII 的降解。

P_{RM}是用来维持溶源态的启动子。它是在从 P_{RE} 开始转录发生之后才发挥作用，这一转录使阻遏物 CI 迅速合成，形成溶源态。这些阻遏物结合到 O_R1 和 O_L2 上，但保持 O_R3 敞开。RNA 聚合酶结合到 P_{RM}上，而 P_{RM}与 O_R3 重叠使结合到 O_R2 上的阻遏物与聚合酶相接。这种蛋白质-蛋白质间的相互作用使得启动子更加有效地运行。高浓度的阻遏物会抑制从 P_{RM}开始的转录，而这一过程可能涉及阻遏物二聚体间的相互作用，即阻遏物二聚体通过 DNA 环实现 O_R1、O_R2 和 O_R3 与 O_L1、O_L2 和 O_L3 的结合。

λ 阻遏物与 RNA 聚合酶间重要的相互作用涉及聚合酶 σ 亚基的 4 区。这条多肽结合在 P_{RM}上弱的 −35 框不远处，使 σ 亚基 4 区与结合到 O_R2 上的阻遏物靠近。因此阻遏物可与 σ 因子相互作用，帮助 RNA 聚合酶结合到弱的启动子上。通过这种方式，O_R2 充当激活位点，而 λ 阻遏物是从 P_{RM}转录的激活子。这样就促使闭合的启动子复合物变成开放的启动子复合物。

某一特定的细菌当被 λ 噬菌体感染后是裂解还是溶源依赖于 *cI* 和 *cro* 基因产物生成速率的竞争。*cI* 基因编码阻遏物，封闭 O_R1、O_R2、O_L1 及 O_L2，同时关闭所有早期基因包括 *cro* 基因的转录，这样就导致溶源。另一方面 *cro* 基因编码 CRO 蛋白来封闭 O_R3 和 O_L3 从而关闭 *cI* 基因的转录，这样就致使细菌裂解。哪一个基因产物首先达到足够高的浓度来阻止其竞争者的合成，它就会在竞争中胜出从而决定细菌的命运。竞争的胜出由 CII 的浓度决定，而 CII 的浓度由细胞内蛋白酶的浓度决定，蛋白酶的浓度又相应地由环境因素如培养基的营养度决定。

当溶源菌遭受 DNA 损伤后，会诱导 SOS 响应。这一反应最初是 RecA 蛋白出现辅蛋白酶活性，使阻遏物将其自身一切两半，从 λ 操纵子上释放出来最终诱导裂解周期。子代 λ 噬菌体以这种方式逃避宿主内的致死性损伤。

复习题

1. 用模型解释噬菌体 SPO1 如何控制转录过程。

2. 总结问题 1 的证据。

3. 描述枯草芽孢杆菌 σ^E 识别 0.4kb 产孢特异性启动子，而 σ^A 识别营养生长特异性启动子的实验证据。

4. 总结枯草芽孢杆菌在孢子发生时调节转录过程的机制。

5. 描述证明枯草芽孢杆菌 σ^E 识别 *spoIID* 启动子，而其他的 σ 因子不能识别 *spoIID* 启动子的实验证据。

6. 解释 *E. coli* 如何对热激做出迅速响应。

7. 用模型解释 T7 噬菌体如何控制转录过程。

8. 在裂解复制过程中 λ 噬菌体如何将转录从早期转换到晚早期再到晚期。

9. 用模型解释 λ 噬菌体感染的 *E. coli* 细胞中 N-介导的终止作用。哪些蛋白质参与了这个过程？

10. 用模型解释 N 和 NusA 在转录终止和抗终止中的作用。

11. 描述证明 N 和 NusA 在内部终止子中控制发夹形成的实验证据。

12. P_{RE}启动子很难被识别并且其本身对 RNA 聚合酶也没有亲和力，那么 *cI* 基因是如何从这个启动子转录的？

13. 如何才能使 CII 和 RNA 聚合酶可以同时结合在 DNA 的同一个位点上？

14. 描述 λ 阻遏物正向和反向调节其自身合成的实验证据。相同的实验表明阻遏物对 *cro* 转录具有怎么样的影响？

15. 图示利用基因间抑制分析来测定两个蛋白质之间相互作用的原理。

16. 描述研究 λ 阻遏物与 *E. coli* RNA 聚合酶 σ 亚基相互作用的实验证据。

17. 用模型解释在λ噬菌体感染的 *E. coli* 中 *cI* 和 *cro* 的竞争是如何导致溶源态或裂解态的建立的。

18. 用模型解释突变诱导λ溶源菌的作用原理。

分析题

1. 有一个基因，怀疑它有两个启动子分别受两个σ因子识别，设计一个实验来证明这一假设。

2. 如果用同一个株系的λ噬菌体感染溶源态 *E. coli* 会发生什么情况？能否得到超感染菌株？为什么？

3. 重复问题2的实验，但使用不同株系的λ噬菌体，其操纵子序列显著不同于溶源态的λ噬菌体。你期望的结果是什么？为什么？

4. 用野生型λ噬菌体感染两个遗传背景不同的 *E. coli* 并分离菌株，分别从它们的噬菌斑中挑选溶源菌，并进行培养。用紫外线照射后，其中一菌株呈裂解态而另一菌株什么也没有得到，分析原因。

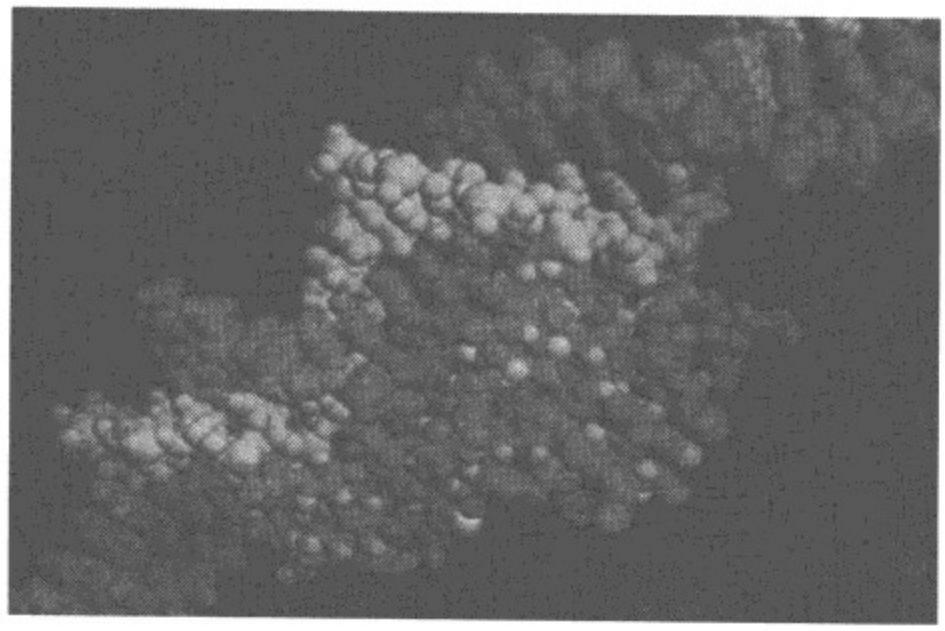
5. 在100%引起裂解态的λ噬菌体中，什么可能发生了突变？

6. 在100%引起溶源态的λ噬菌体中，什么可能发生了突变？

7. 如果实验中所用的λ噬菌体的 *N* 基因失活，期望得到的细菌是什么样子？是裂解态的、溶源态的、两种都有、两种都没有。为什么？

翻译　刘智捷　校译　郑用琏

第9章　细菌中DNA-蛋白质的相互作用



基因调控：Cro蛋白结合到DNA上的计算机模型

© *Ken Eward/SS/Photo Researchers，Inc.*

在第7章和第8章我们讨论了与DNA特定位点紧密结合的几种蛋白质，包括RNA聚合酶、*lac*阻遏物、CAP、*trp*阻遏物、λ阻遏物以及Cro。对这些蛋白质已有详细的研究，它们都能在大量的无关序列中定位并结合到一段特殊的短DNA序列上。那么，这些蛋白质如何完成这种大海捞针似的专一性结合的呢？除了RNA聚合酶外，它们都含有相似的结构基序，即由一个短的蛋白质转角所连接的两个α螺旋。这种螺旋-转角-螺旋基序（图9.1a）使第二个螺旋（识别螺旋）可以合适地嵌入目标DNA的大沟里（图9.1b）。我们将看到这种巧妙的匹配构象在不同蛋白质间有很大变化，但是都能像钥匙与锁那样准确契合。在本章，我们将学习几个研究得比较透彻的例子，了解原核细胞中DNA-蛋白质的相互作用，搞清楚什么东西使它们如此专一。在第12章我们将讨论真核生物中的另外几种DNA结合基序。

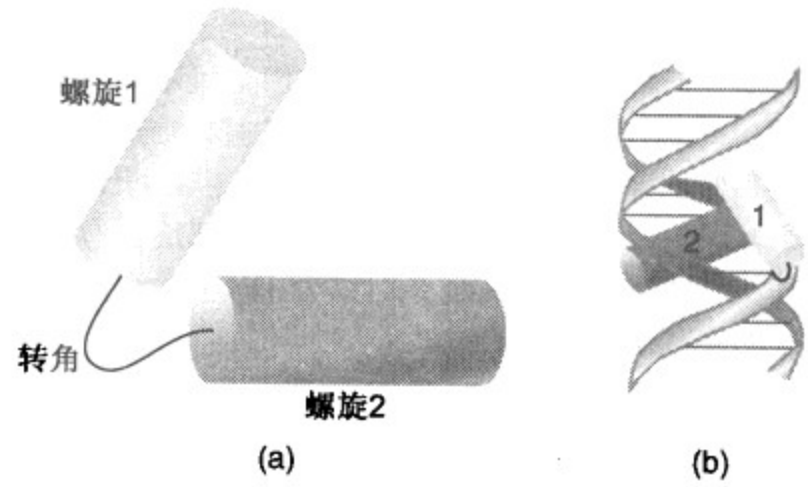


图9.1　螺旋-转角-螺旋基序是DNA结合的元素。(a) λ阻遏物的螺旋-转角-螺旋基序；(b) 阻遏物单体的螺旋-转角-螺旋基序对λ操纵基因的配合。基序的螺旋2插入目标DNA的大沟内，其背面（离开读者）的一些氨基酸可与DNA结合。

9.1　λ阻遏物家族

如图9.2所示，λ及类似噬菌体的阻遏物识别螺旋位于相应操纵基因DNA的大沟内，这种结合的特异性依赖于识别螺旋的某些氨基酸与伸入DNA大沟的某些碱基的功能基团的特异性结合，以及与DNA骨架上的磷酸基团的特异性结合。其他具有螺旋-转角-螺旋基序的蛋白质因为识别螺旋缺乏正确的氨基酸，而不能与同样的DNA位点很好地结合。我们想知道哪些氨基酸在这种结合中起重要作用。

利用定点突变探测结合的专一性

Mark Ptashne及同事利用其对两类λ噬菌体**434**和**P22**及其操纵基因的研究提供了专一性结合的部分答案。这两个噬菌体的分子遗传特性非常

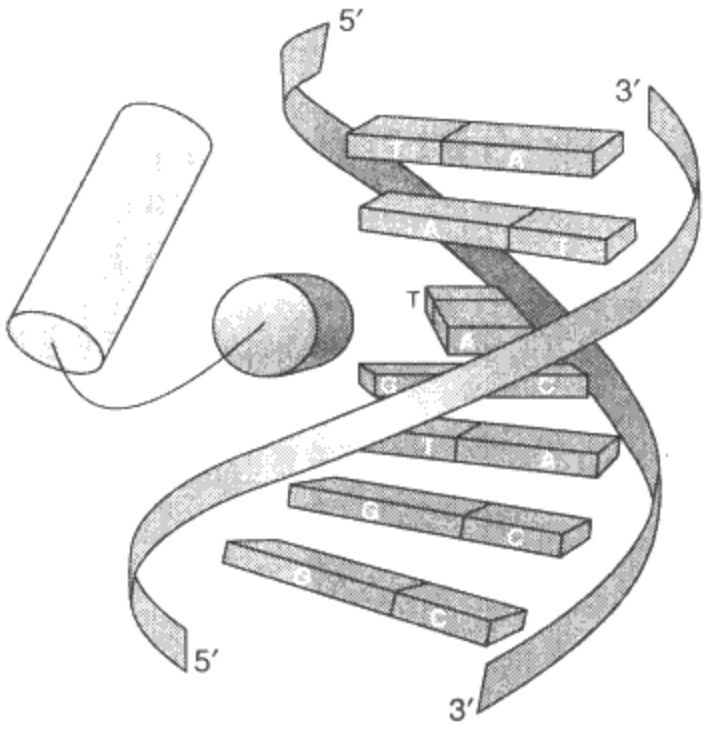


图 9.2　λ 阻遏物单体的识别螺旋与 DNA 操纵基因大沟相配合的示意图。识别螺旋用红色圆柱体表示，在大沟中处于有利于结合 DNA 碱基对边缘氢键的位置。(*Source*: Adapted from Jordan, S. R. and C. O. Pabo. Structure of the lambda complex at 2.5 Å resolution. Details of the repressor-operator interaction. *Science* 242: 896, 1988.)

类似，但具有不同的免疫区，它们合成不同的阻遏物，识别不同的操纵基因。两种阻遏物都与 λ 阻遏物类似，含有螺旋-转角-螺旋基序，但由于识别操纵基因的碱基序列不同，推测其识别螺旋具有不同的氨基酸组成，尤其是专门结合 DNA 大沟里的碱基的氨基酸。

Stephen Harrison 和 Ptashne 利用 X 射线衍射技术（知识窗 9.1）分析操纵基因-阻遏物复合体，鉴别出 434 噬菌体阻遏物的识别螺旋与操纵基因大沟里的碱基的结合面。通过类比分析，对 P22 噬菌体的阻遏物也做出了相似的预测，图 9.3 示意性地显示了每个阻遏物中最有可能与操纵基因结合的氨基酸。

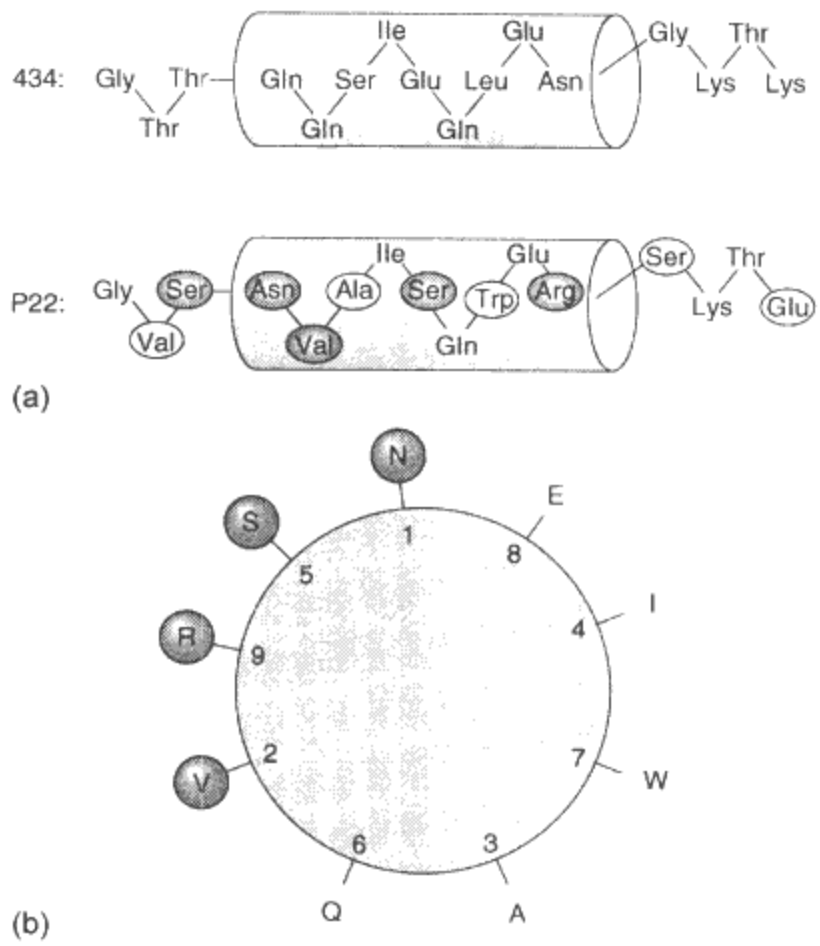


图 9.3　两个类 λ 噬菌体阻遏物的识别螺旋。(a) 两个阻遏物识别螺旋的关键氨基酸。显示 434 和 P22 阻遏物识别螺旋的氨基酸序列以及两侧的几个氨基酸。圆圈中的氨基酸为两个蛋白质中不相同的，它们很可能在特异性方面起作用。螺旋面向 DNA 的氨基酸很可能参与结合 DNA。改变这些氨基酸以及螺旋之前的一个氨基酸可以改变对 DNA 结合的特异性。(b) P22 阻遏物的识别螺旋端面观。数字代表蛋白链中的氨基酸位置。螺旋的左手侧朝向 DNA，该侧的氨基酸在结合方面很重要。434 阻遏物相应位置上的不同氨基酸用红圈标出。(*Source*: (b) Adapted from Wharton, R. P. and M. Ptashne, Changing the binding specificity of a repressor by redesigning an alpha-helix. *Nature* 316: 602, 1985.)

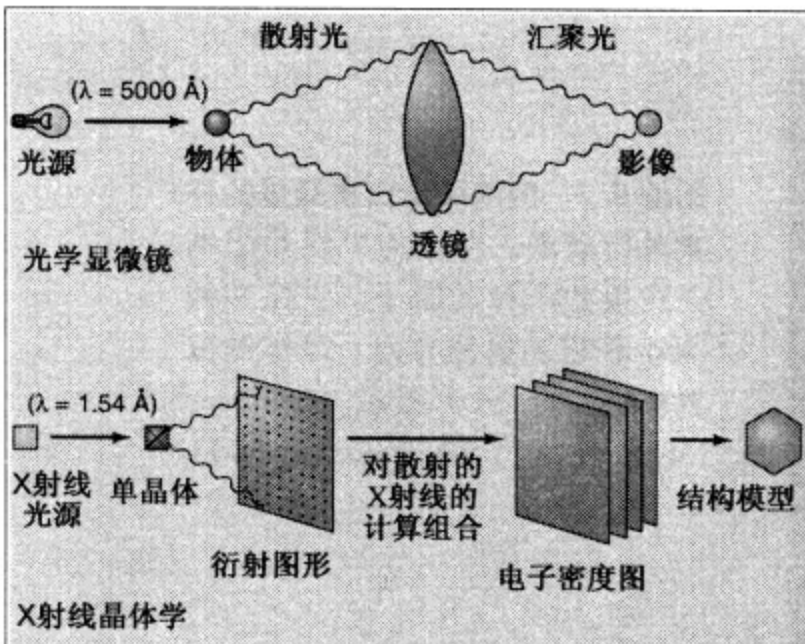
知识窗 9.1

X 射线晶体学

本书有很多 DNA 结合蛋白结构的例子，这些结构是通过 **X 射线衍射分析**（X-ray diffraction analysis）获得的，该方法也称为 **X 射线晶体学**（X-ray crystallography）。此处将对这种非常有用的技术做一介绍。

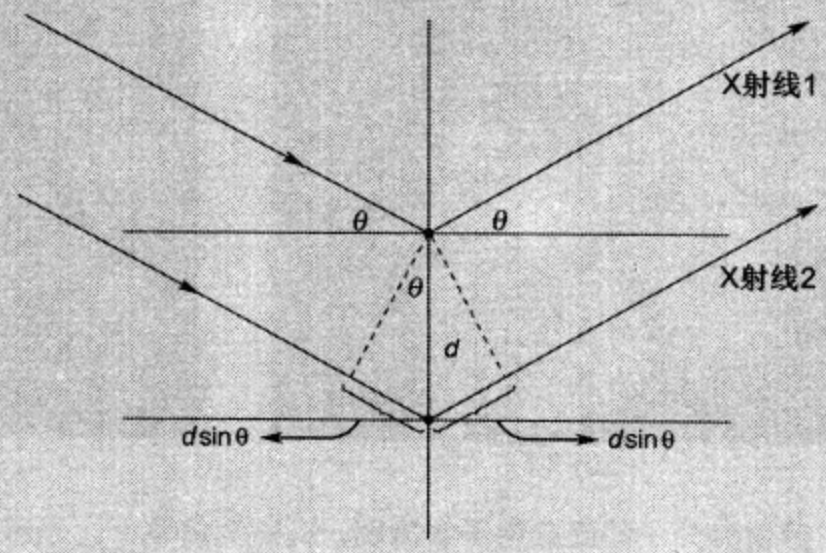
就像光线一样，X 射线是电磁辐射，但其波长短得多，所以能量更高。因此，X 射线衍射分析的原理在某些方面与光学显微镜相似就不足为奇了。图框 9.1 显示了这种相似性。在光学显微镜中（图框 9.1a），可见光被物体散射，透镜收集光线，聚焦形成该物体的影像。

在 X 射线衍射中，X 射线被物体（晶体）散射。但我们遇到了一个重要问题，即没有可聚焦 X 射线的透镜，所以必须使用相关的间接方法成像。该方法基于以下的条件：当 X 射线和原子周围的电子云相互作用时，X 射线向任一方向散射。然而，由于 X 射线束与很多原子作用，大部分散射的X射线由于其



图框 9.1　光学显微镜（上）和 X 射线晶体成像（下）中图像重建过程示意图。

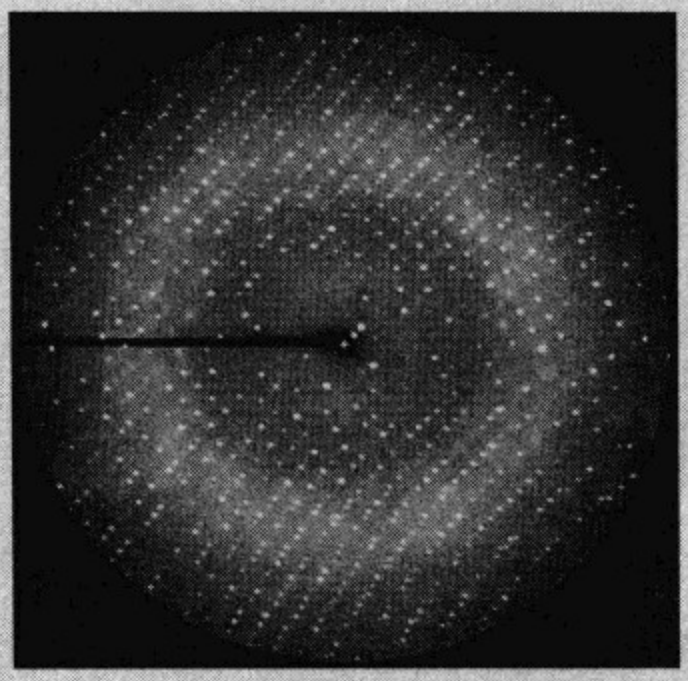
波动性而彼此抵消。但是散射到特定方向的 X 射线会被衍射现象所增强。Bragg 定律 $2d\sin\theta=\lambda$，描述了衍射角（θ）和散射平面间距（d）的关系。如图框 9.2 所示，X 射线 2 经过的距离比 X 射线 1 长 $2\times d\sin\theta$。因此，如果 X 射线 2 的波长（λ）等于 $2d\sin\theta$，则 X 射线 1 和 X 射线 2 散射后因相位相同而得到增强。如果 λ 不等于 $2d\sin\theta$，散射后的射线将被减弱。这些衍射的 X 射线被放置在 X 射线路径上的收集装置（监测器）记录为点子。该装置可以简单到只是一张 X 射线感光胶片，现在已有更高效的电子监测器。图框 9.3 是一个简单蛋白溶菌酶的衍射图。尽管该蛋白相对简单（只有 129 个氨基酸），由点组成的衍射图却很复杂。为了获得蛋白质的三维结构，必须旋转晶体，记录不同方位的衍射图。



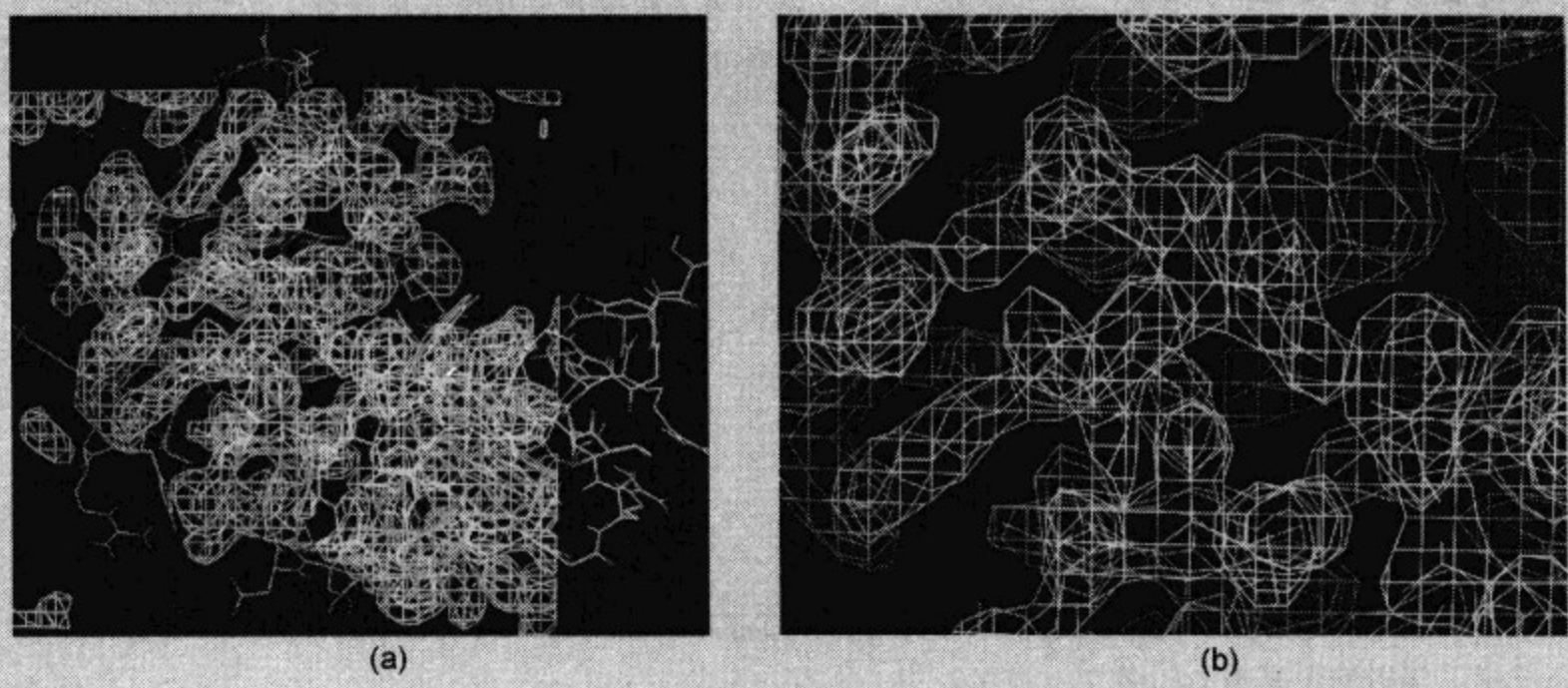
图框 9.2　通过晶体平行平面的两束 X 射线的反射。两条 X 射线（1 和 2）以夹角 θ 射向晶体平面，并以相同角度反射。两平面间的距离是 d。X 射线 2 额外辐射的距离是 $2d\sin\theta$。

下一个任务是利用衍射图中的点阵绘出产生衍射的分子结构。不幸的是不能用衍射图的点阵重建电子**密度图**（electron-density map）（电子云分布），因为散射图不含单反射相角（phase angle）的物理参数信息。为解决这个问题，晶体学家用重金属溶液（Hg、Pt、U 等）浸润蛋白晶体制成含 3～10 种不同重金属的衍生晶体。这些重金属原子倾向于结合有反应活性的氨基酸残基，例如半胱氨酸、组氨酸、天冬氨酸，但不改变蛋白质的结构。这种方法叫做多重同形置换（multiple isomorphous replacement，MIR）。通过对比原始的衍射图和重金属原子衍生晶体的衍射图，可以确定单反射相角。一旦得到相角，就可以精确地把衍射图通过数学计算转换为该衍射分子的电子密度图。然后就可用电子密度图推断该衍射分子的结构了。用衍射射线建立的衍射物体图像与用透镜的相似，但不是像透镜那样获得物理上的图像，而是通过数学计算获得的。图框 9.4 是溶菌酶部分结构的电子密度图，杆状图是从密度图推断出的分子结构。图框 9.5 是根据整个分子的电子密度图推断出的完整溶菌酶分子的 3 种不同表示形式。

为什么在 X 射线衍射分析中要使用单晶体呢？将单一蛋白质分子置于 X 射线路径上显然是不实际的。就算我们能做到，单一分子衍射的能量也太弱了难以检测。因此，需要在 X 射线束中放置很多蛋白质分子，产生足够强的信号以便检测。为什么不用蛋白质粉末或者蛋白质溶液呢？问题是粉末或溶液中的分子是随机定向的，样品衍射出的 X 射线无法解析。



图框 9.3　溶菌酶蛋白质晶体的样本单衍射图。左边的黑线是阻挡射线束的小针的影子，它保护探测器不受 X 射线直射。晶体的位置由中心的（+）指示。（*Source*：Courtesy of Fusao Takusagawa.）

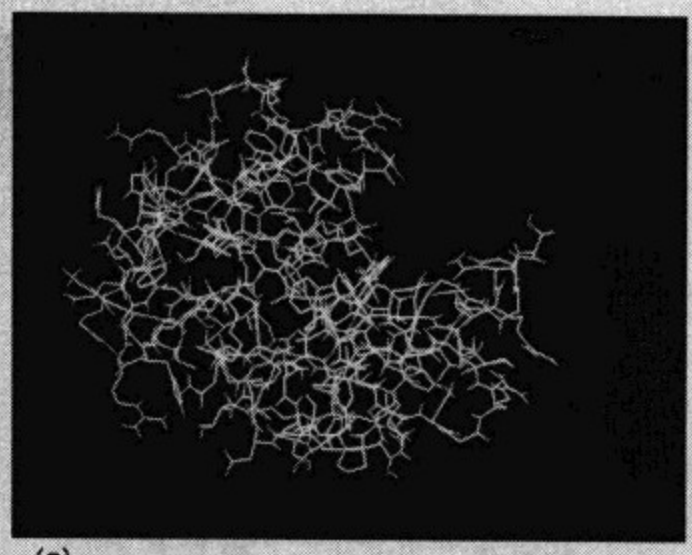


(a)　　(b)

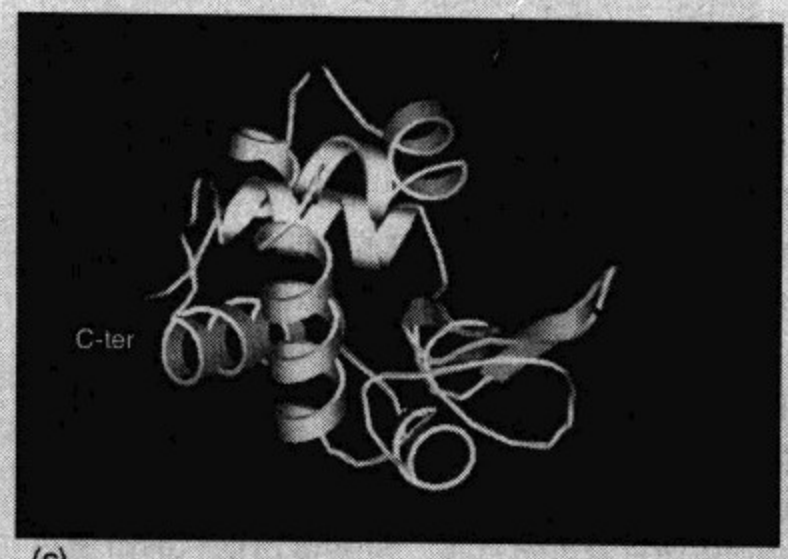
图框 9.4　溶菌酶分子的部分电子密度图。(a) 低倍镜下，显示大部分分子的电子密度图。蓝色部分对应高电子密度区域。从电子密度图推断它们聚集成该分子的棒状模型（红、黄和蓝）。(b) 高倍镜下，显示 (a) 图中的中心部分。分辨率是 2.4Å，无法分辨出单个原子，但足以确定每个氨基酸的独特形态。（*Source*：Courtesy Fusao Takusagawa.）

解决这个问题的办法是用蛋白质晶体。晶体由很多在三维空间中规则排列的小重复单位（晶胞，unit cell）组成。蛋白质的一个晶胞含有几个分子，它们通常是空间对称性相关的。因此，晶体中晶胞内所有分子的衍射相同，而且彼此增强。要使 X 射线衍射结果可用，蛋白质晶体的最小尺度至少应 0.1mm。这种大小的立方晶体含有 10^{12} 个以上分子（假设一个蛋白质分子占据 50Å×50Å×50Å 的空间）。图框 9.6 是适合做 X 射线衍射分析的溶菌酶晶体照片。蛋白质晶体不只含有纯蛋白质，也含有大量的溶剂（占晶体重量的 30%～70%）。因此，蛋白在晶体中的环境类似于在溶液中，晶体的三维结构与溶液中的相近。总之，我们确信通过 X 射线晶体成像确定的蛋白质结构与它们在细胞中的结构相近。事实上，多数酶晶体都保持了酶活性。

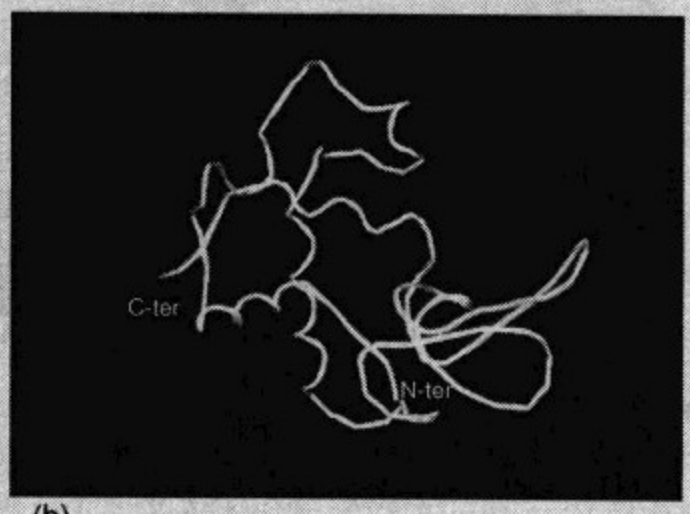
为什么不用可见光观察蛋白质结构避免 X 射线的诸多麻烦呢？问题在于**分辨率**（resolution），即区别分子不同部位的能力。分析分子结构的最终目的是分辨出每个原子，使分子中所有原子间的精确空间关系能够呈现。但是原子的空间尺度是埃（Å）级的，而射线的最大分辨力是其波长的 $1/3(0.6\lambda/2\sin\theta)$。所以我们需要波长（以埃为单位测量的）很短的射线来分辨蛋白质中的原子。可见光的平均波长约为 500nm(5000Å)。用可见分辨出原子显然是不可能的。而 X 射线的波长为 1 到几个埃。例如，被激发的铜原子发射出的特征 X 射线波长是 1.54Å，对蛋白质的高分辨率 X 射线衍射分析是理想的。



(a)



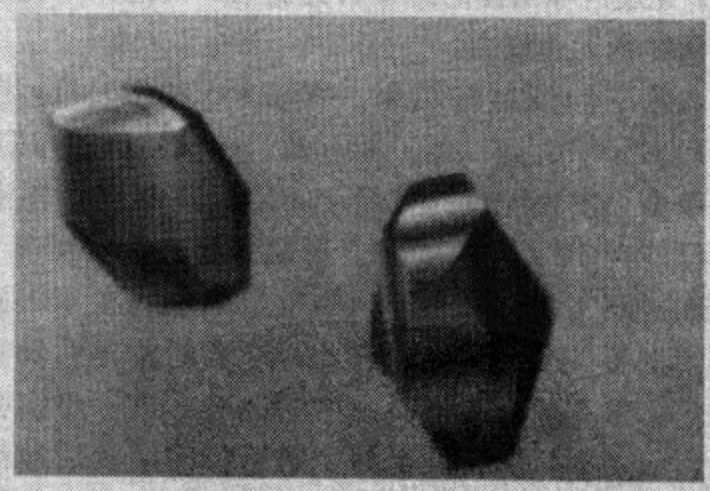
(c)



(b)

图框 9.5　根据电子密度图框 9.4 计算出的溶菌酶结构的 3 种形式。（a）如图框 9.4 的棒状图。（b）线状图。α 螺旋为绿色，β 折叠片为红色，随机卷曲为蓝色。蛋白质的 N 端和 C 端分别标记为 N-ter 和 C-ter。（c）带状图。颜色标记和（b）中相同。α 螺旋的螺旋性质在图中显而易见。3 个图中都有的右上部裂缝是酶的活性位点。（*Source*：Courtesy Fusao Takusagawa.）

在本章中我们会看到不同分辨率水平的蛋白质结构。分辨率不同是什么原因造成的呢？在蛋白质晶体中，蛋白质分子相对有序的排列可以在远离光束的地方产生很多衍射点，即远离监测器中心。这些点是由大衍射角（θ，图框 9.2）的 X 射线产生的。根据有序晶体的衍射点计算出的电子密度图可以展示出衍射分子的高分辨率图像。另一方面，分子排列不太规则的蛋白质晶体得出的衍射点只靠近检测器中心，是由小衍射角的 X 射线产生。这种数据产生的分子图像分辨率相对较低。



图框 9.6　溶菌酶晶体。用偏振滤光片拍摄以便产生有颜色的晶体照片。晶体的实际大小接近 0.5mm×0.5mm×0.5mm。（*Source*：Courtesy Fusao Takusagawa.）

分辨率和衍射角之间的关系是 Bragg 定律 $2d\sin\theta=\lambda$ 的另一个结果。变换 Bragg 等式，$d=\lambda/2\sin\theta$，可见 d（蛋白质中结构元件之间的距离）和 $\sin\theta$ 呈反比。因此，晶体中结构元件之间的距离越大，衍射角越小，越靠近衍射图的中心。即低分辨率结构（即元件之间的距离大）产生的点子出现在衍射图的中间部分。同样，高分辨率结构产生的点子出现在图的外围，因为这种晶体以大角度衍射 X 射线。当晶体学家制备出足够好的能产生高分辨率的晶体时，他们就能建立该蛋白质结构的精细模型。

我们在本章讨论的蛋白质是 DNA 结合蛋白。在多数情况下，研究者已经制备了蛋白质和含有该蛋白质识别序列的双链 DNA 片段的共晶体。这让我们不仅能知道该蛋白质和 DNA 在蛋白质-DNA 复合体中的形状，也能知道在蛋白质-DNA 相互作用中起作用的原子。

如果这些氨基酸确实很重要，那么只要对其加以改变就可改变阻遏物的特异性。可以改变 434 阻遏物使它不识别自身操纵基因而识别 P22 操纵基因。这正是 Robin Wharton 和 Ptashne 他们所做的。首先克隆得到 434 阻遏物基因，然后利用第 5 章的基因突变技术系统地将 434 阻遏物识别螺旋的 5 个氨基酸密码子突变成 P22 阻遏物识别螺旋的 5 个相应氨基酸。

然后，在细菌中表达这些突变基因，并通过体内和体外实验检测基因产物结合 434 及 P22 操纵基因的能力。体内实验检测免疫性。回想一下，感染 λ 噬菌体的 *E. coli* 细胞进入溶源状态后便对 λ 噬菌体的再次感染具有免疫力，溶源菌中额外的阻遏物会立刻与入侵的 λ DNA 结合并抑制其表达（第 8 章）。434 和 P22 噬菌体是类 λ 噬菌体（lambdoid），但它们的免疫区以及操纵基因是有区别的。因此，434 溶源菌对 434 噬菌体再次感染有免疫力但对 P22 再次感染无免疫力，反之亦然，P22 溶源菌对 P22 噬菌体再次感染有免疫力而对 434 噬菌体则没有。

Wharton 和 Ptashne 没有培养溶源菌，而是利用编码重组 434 阻遏物的质粒转化 *E. coli*，然后检测重组 434 阻遏物（其识别螺旋已被改变成 P22 的识别螺旋）是否仍具有原来的结合特异性。如果有，则转化了重组阻遏物的细胞对 434 的感染就应该有免疫力，如果其结合特异性发生变化则应该对 P22 的感染有免疫力。实际上 434 和 P22 都不感染 *E. coli*，所以研究者用 434 和 P22 免疫区（λ_{imm}434 和 λ_{imm}P22）的重组 λ 噬菌体进行实验，发现突变 434 阻遏物的细菌对 P22 免疫区噬菌体的感染有免疫力，而对 434 免疫区噬菌体则无免疫力。

为了验证这些结果，Wharton 和 Ptashne 在体外用 DNase 足迹实验（第 5 章）检测阻遏物与 DNA 的结合，发现纯化的重组 434 阻遏物与 P22 阻遏物一样能在 P22 操纵基因上形成“足迹”（图 9.4）。而在对照实验中（未给出）重组阻遏物在 434 操纵基因上不再形成足迹。因此，结合的特异性确实因这 5 个氨基酸的改变而发生了改变。进一步实验表明，前 4 个氨基酸对于两个阻遏物的结合活性都是充分必要的。如果阻遏物识别螺旋有 TQQE（苏氨酸、谷氨酰胺、谷氨酰胺、谷氨酸），阻遏物就与 434 操纵基因结合，如果是 SNVS（丝氨酸、天冬氨酸、缬氨酸、丝氨酸）则与 P22 操纵基因结合。

如果 Wharton 和 Ptashne 不是改变而是消除阻遏物的特异性，结果会如何呢？他们可由此鉴定出阻遏物中决定特异性的重要氨基酸，然后对其进行随机突变，以显示重组 434 阻遏物不再结合其操纵基因。如果这正是他们所做的全部内容，则可认为实验结果与所提出的假设一致，即这些改变的氨基

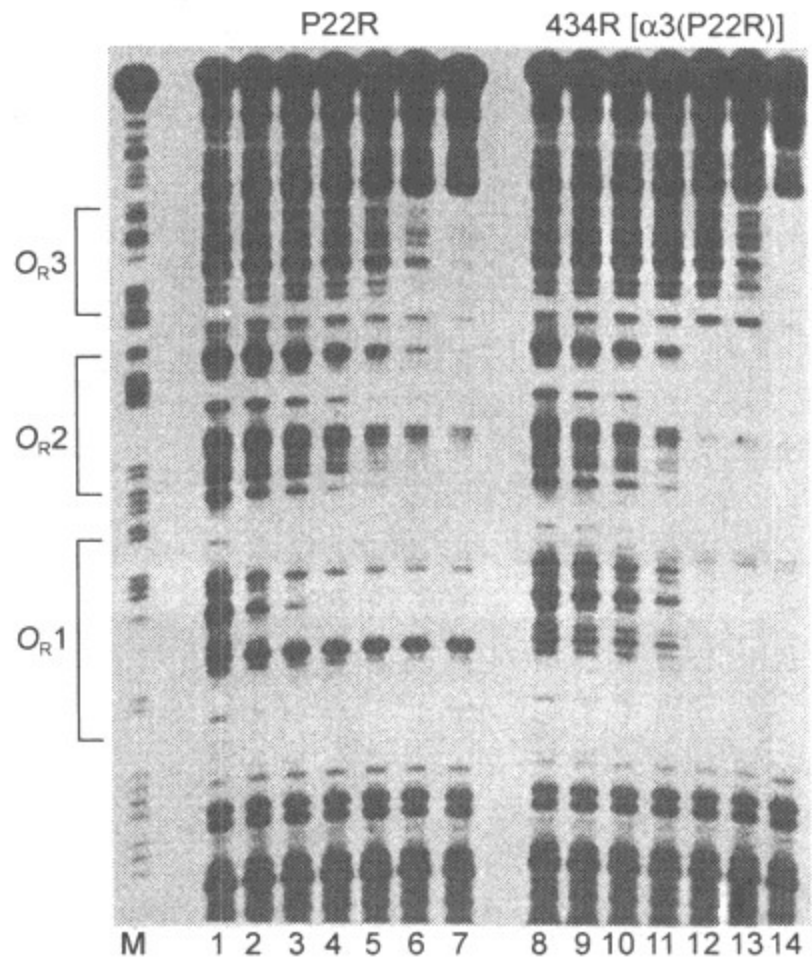


图 9.4　重组 434 阻遏物的 DNase 足迹。Wharton 和 Ptashne 用末端标记的 P22 噬菌体的 O_R 与不同阻遏物做 DNA 酶足迹实验。阻遏物分别是 P22 阻遏物（P22，泳道 1～7）或 434 阻遏物其识别螺旋（α 螺旋 3）的 5 个氨基酸被突变成 P22 识别螺旋（434R［α3（22）］，泳道 8～14）的相应氨基酸。两组泳道各阻遏物的浓度都逐渐增加（泳道 1 和 8 为 0mol/L，泳道 2～7 为 7.6×10^{-10}～1.1×10^{-8} mol/L，泳道 8～14 为 5.2×10^{-9}～5.6×10^{-7} mol/L）。Marker 泳道是测序产物的A+G反应。所有 3 个右向操纵基因的位置在左边的括号中标出。（*Source*：Wharton，R. P. and M. Ptashne，Changing the binding specificity of a repressor by redesigning an alpha-helix. *Nature* 316（15 Aug 1985），f. 3，p. 603. © Macmillan Magazines Ltd.）

酸直接参与结合。但也有另一种解释，认为这些氨基酸只是对阻遏物的三维构象重要，改变它们就改变了蛋白质的构象，从而间接地抑制了结合。相比之下，通过改变氨基酸而改变结合特异性有力地证明了这几个氨基酸直接参与结合。

Ptashne 小组在相关的 X 射线晶体研究中证明，λ 阻遏物有一个氨基末端臂，但在 434 和 P22 噬菌体中未发现。该臂通过环抱操纵基因而使阻遏物结合 λ 操纵基因。图 9.5 是 λ 阻遏物二聚体与 λ 操纵基因结合的计算机模型。可以看到，顶部单体的螺旋-转角-螺旋基序伸向 DNA 大沟中，底部单体的臂

接近并环抱 DNA。

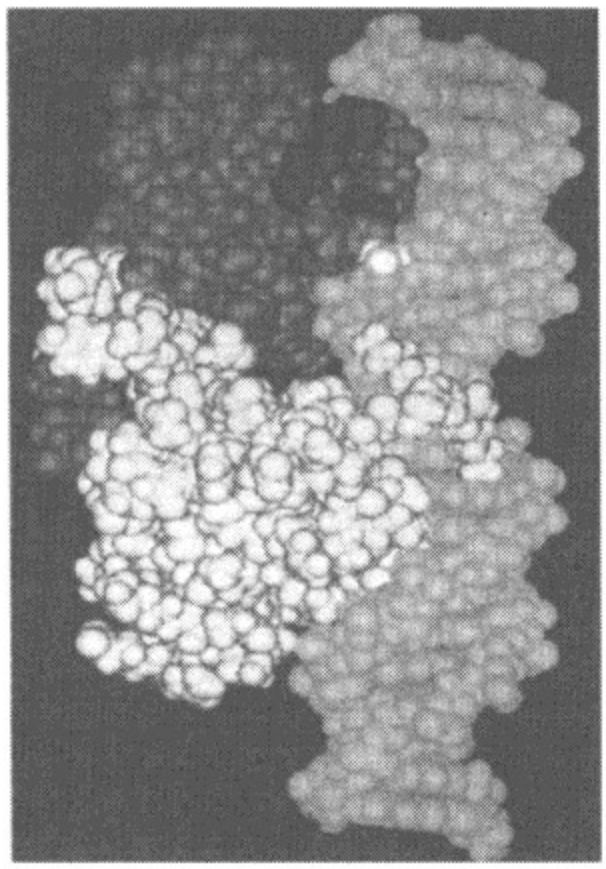


图 9.5 λ 阻遏物二聚体结合 λ 操纵基因（O_R2）的计算机模型。DNA 双螺旋在右侧。阻遏物的两个单体为深蓝色和黄色。上方单体的螺旋-转角-螺旋基序（红色和蓝色）插入 DNA 的大沟。下方单体的臂环抱 DNA。（*Source*: Hochschild, A., N. Irwin, and M. Ptashne, Repressor structure and the mechanism of positive control. *Cell* 32 (1983) p. 322. Reprinted by permission of Elsevier Science. Photo by Richard Feldman.）

Cro 也以螺旋-转角-螺旋基序结合与 λ 阻遏物同样的操纵基因，但是对这 3 个操纵基因（见第 8 章）的亲和性与 λ 阻遏物刚好相反，先结合 O_R3，最后是 O_R1。因此，通过改变识别螺旋的氨基酸可以辨别出使 Cro 和 λ 阻遏物具有不同识别特异性的氨基酸。Ptashne 小组完成了这一研究，发现与 λ 阻遏物的氨基末端臂一样，识别螺旋的第 5、6 位氨基酸对结合的专一性特别重要。通过改变操纵基因的碱基，发现识别 O_R1 和 O_R3 的关键碱基对在 3 号位，Cro 对其更敏感，而 5 号和 8 号位对阻遏物结合是选择性的。

小结　类 λ 噬菌体阻遏物都具有与操纵基因 DNA 大沟匹配结合的识别螺旋。识别螺旋中面向 DNA 的某些氨基酸特异性地结合操纵基因的碱基，这种结合决定了蛋白质-DNA 相互作用的特异性。改变这些氨基酸可改变阻遏物的特异性。λ 阻遏物自身有一个额外氨基末端臂，在其他阻遏物中未发现。氨基末端臂通过环抱而帮助结合 DNA。λ 阻遏物和 Cro 蛋白对相同操纵基因的亲和性一致，但它们对 O_R1 或 O_R3 具有细微差别，这种差别取决于其识别螺旋中不同氨基酸与两个操纵基因中不同碱基对的相互作用。

高分辨率分析 λ 阻遏物-操纵基因相互作用

Steven Jordan 和 Carl Pabo 希望以更高的分辨率来观察 λ 阻遏物-操纵基因的相互作用。通过对阻遏物片段与操纵基因片段的高质量共结晶，他们得到了 2.5Å 的分辨率。阻遏物片段有 1～92 位残基，包括全部 DNA 结合域，操纵基因片段（图 9.6）20bp 长，含有阻遏物二聚体的完整结合位点，即含有两个分别与阻遏物的单体结合的半位点（two half-sites）。X 射线晶体衍射中常用部分分子做结晶，得到比全蛋白质或全长 DNA 更好的晶体。由于该实验的主要目的是阐明阻遏物与操纵基因结合面的结构，所以 DNA 片段和部分蛋白质可能与完整蛋白质和 DNA 一样好用，因为它们都含有所需的元件。

```
    1 2 3 4 5 6 7 8 9
T A T A T C A C C G C C A G T G G T A T
  T A T A G T G G C G G T C A C C A T A A
                    8′ 7′ 6′ 5′ 4′ 3′ 2′ 1′
```

图 9.6　用于制备操纵基因-阻遏物共结晶的操纵基因片段。这个 20mer 片段含有两个 λO_L1 半位点，各与阻遏物的一个单体结合。半位点包含在黑体表示的 17bp 区内，各有 8bp，被中间的 G-C 对隔开。左边的半位点有一个共同序列，右边的与共同序列有点偏差。共同序列半位点的碱基对标以数字 1～8；另一半标以数字 1′～8′。

一般结构特征　本章介绍的 λ 阻遏物与操纵基因精确配合的图示是依据 Steven Jordan 和 Carl Pabo 的高分辨率模型绘制的。图 9.7 是同一模型的更为详细的图示，以揭示蛋白质-DNA 相互作用的几个主要方面。首先，每个阻遏物单体的识别螺旋（3 和 3′，红色）以两个半位点嵌入在 DNA 大沟里。还可看到螺旋 5 和 5′相互靠近将阻遏物二聚体拉在一起。最后，注意 DNA 的形状与标准 B 型 DNA 类似，还可看到 DNA 有稍许弯曲，尤其在其两端曲向阻遏物二聚体，但螺旋的其余部分是相对直的。

与碱基的相互作用　图 9.8 详细地描述了阻遏物单体的氨基酸与半位点中操纵基因碱基的相互作用。参与相互作用的关键氨基酸是 Gln33、Gln44、Ser45、Lys4 及 Asn55。图 9.8a 是这种相互作用的立体图，其中 α 螺旋 2 和 3 用粗线表示，识别螺旋（3）几乎与纸面垂直，所以螺旋的多肽骨架看起来像一个环，关键氨基酸与 DNA 碱基形成氢键（虚线标出）。

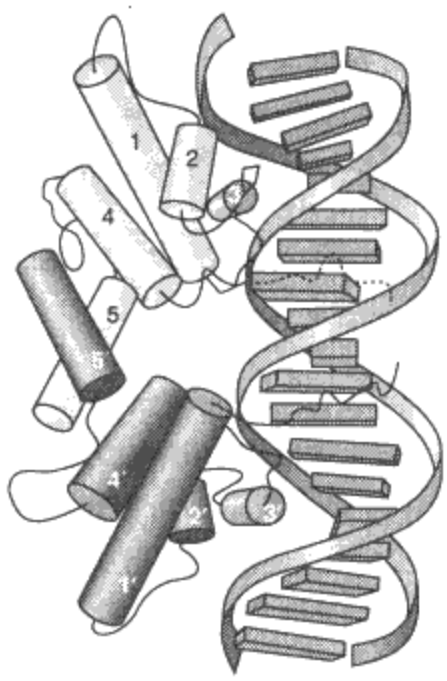


图 9.7　λ 阻遏物-操纵基因复合体的几何学。DNA 被结合到阻遏物二聚体上，每个单体的识别螺旋标为 3 和 3′。（*Source*：Adapted from Jordan，S. R. and C. O. Pabo，Structure of the lambda complex at 2. 5 Å resolution. Details of the repressor-operator interaction. *Science* 242：895，1988.）

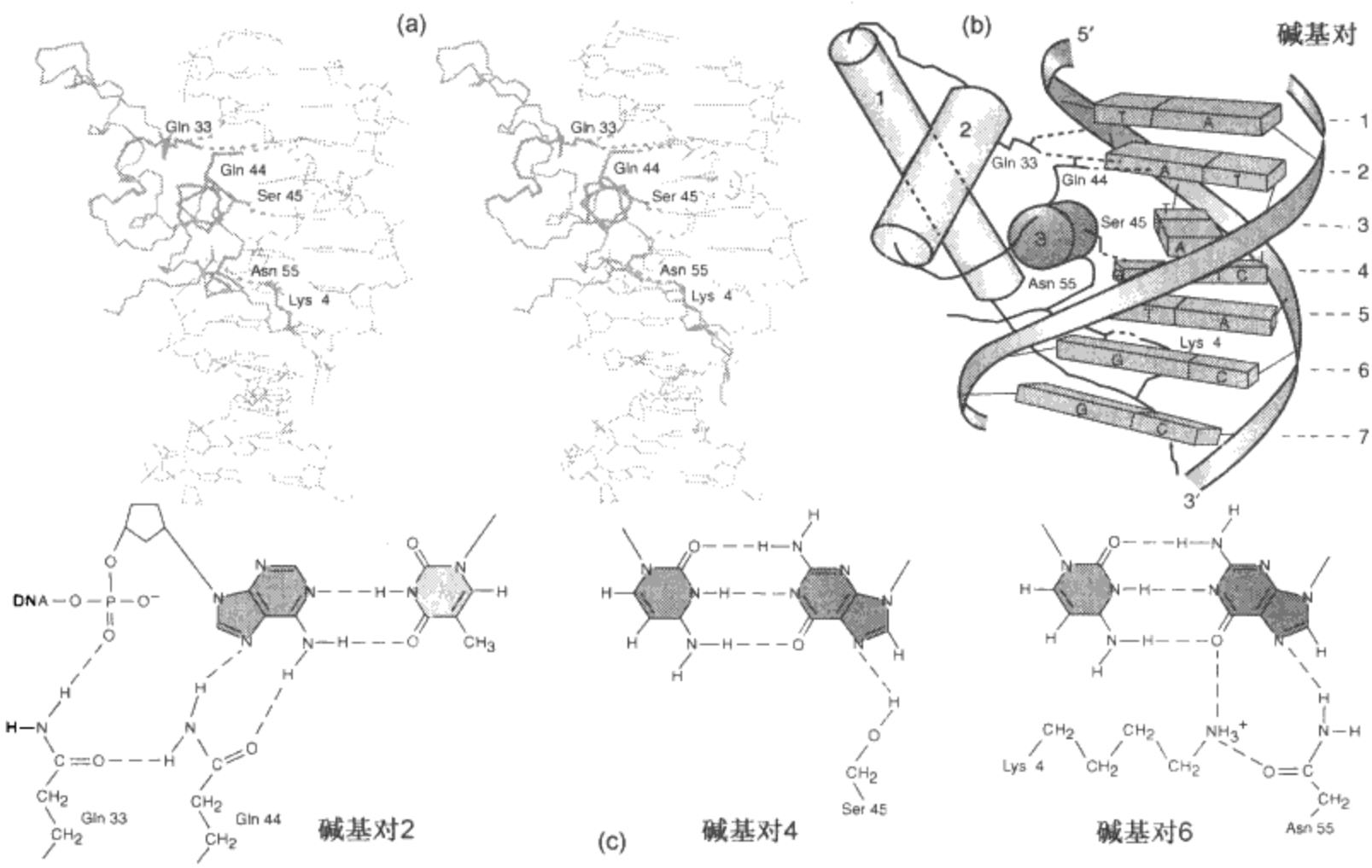


图 9.8　λ 阻遏物与操纵基因大沟里碱基间的氢键。（a）复合体的空间示意图，DNA 双螺旋在右侧，阻遏物单体的氨基末端在左侧。α 螺旋 2 和 3 用粗线表示，识别螺旋几乎与纸面垂直。氢键用虚线表示。（b）（a）中的氢键示意图。只标出重要的氨基酸。碱基对数字编号在右侧。（c）氢键的详细结构。给出了关键氨基酸侧链和碱基结构以及它们所参与形成的氢键。（*Source*：Reprinted with permission from Jordan，S. R. and C. O. Pabo，Structure of the lambda complex at 2. 5 Å resolution：Details of the repressor-operator interactions. *Science* 242：896，1988. Copyright © 1988 AAAS.）

图 9. 8b 是同一氨基酸/DNA 相互作用的示意图，但更容易观察到氢键。可以看到识别螺旋中氨基酸与 DNA 结合的其中 3 个氢键，Gln44 与腺嘌呤 2形成 2 个氢键，Ser45 与鸟嘌呤 4 形成 1 个氢键。图 9. 8c 给出了 a、b 图中这些氢键的详细图示，Gln44 与 Gln33 形成氢键，而 Gln33 又与 2 号碱基的磷酸形成氢键，这是一个氢键网络的例子，它涉及 3 个或更多实体（如氨基酸、碱基或 DNA 骨架），因此 Gln33 的参与十分重要。通过为 DNA 骨架与 Gln44 搭桥，Gln33 使 Gln44 及识别螺旋上的其他部分处于

和操纵基因作用的最佳位置。因此，虽然 Gln33 位于螺旋 2 的开始处而不在识别螺旋上，它却在蛋白质-DNA 相互作用中发挥重要作用。为进一步强调 Gln33 的重要性，请注意在 434 噬菌体阻遏物的相同位置也有这个氨基酸，并且与 434 操纵基因有相同的作用，这一点我们将在下面的章节中看到。

Ser45 碱基对与 4 上的鸟嘌呤碱基形成重要氢键，其亚甲基（CH2）碱基对与 5 上胸腺嘧啶的甲基相互靠近参与疏水作用。疏水作用可能还包括 Ala49，涉及非极性基团如甲基和亚甲基，这些基团聚在一起躲避水溶液的极性环境，就像油滴聚在一起减少与水接触一样。疏水性的字面意思是“惧水”。

与碱基形成氢键的另外两个氨基酸既不属于识别螺旋也不属于任何螺旋。Asn55 位于第 3 和第 4 螺旋的连接处，Lys4 位于绕 DNA 的臂上。在这里我们又看到氢键网络的例子，它们不仅在氨基酸与碱基间形成，而且在两个氨基酸间也有。图 9.8c 非常清楚地显示了两个氨基酸与鸟嘌呤间所形成的氢键以及氨基酸之间的氢键。这种氢键网络使复合体的结构更加稳定。

氨基酸/DNA 骨架的相互作用　我们已经看到 Gln33 与 DNA 骨架（碱基对 1 和 2 之间的磷酸基）形成氢键的例子，但这只是半位点中的 5 个相互作用之一。图 9.9 描述了半位点处的相互作用，其中包括 5 种不同氨基酸的作用，但只有一个氨基酸（Asn52）位于识别螺旋内。虚线代表多肽骨架上 NH 基团所形成的氢键而不是氨基酸侧链所形成的。

其中，Gln33 的 NH 氢键令人感兴趣，它为螺旋 2 提供所有的电子。回忆第 3 章介绍的蛋白质 α 螺旋的所有 C═O 键都指向一个方向，由于每个 C═O 键都是极性的，在氧原子上带部分负电荷，在碳原子上带部分正电荷。螺旋的氨基末端富含正电子使整个螺旋带有相当强的极性，因而带正电的螺旋末端会自然地与 DNA 骨架的负电荷吸附。再回头看看图 9.9，注意 2 号螺旋末端的 Gln33 直接对应 DNA 骨架，使螺旋末端带正电的氨基酸与 DNA 负电荷间的电荷吸附达到最大，稳定 Gln33 的 NH 基团与 DNA 骨架磷酸基团间的氢键。

参与氨基酸侧链与 DNA 骨架磷酸基团间的其他氢键还有 Lys19、Asn52 与磷酸 P_B 间形成的氢键。Lys26 带有完全电荷，尽管与 DNA 骨架相距甚远不能直接与磷酸基团结合，但它有助于蛋白质-DNA 的吸附。众多氨基酸/DNA 磷酸间的结合表明，这些相互作用对稳定蛋白质-DNA 复合体起重要作用。图 9.9 还显示 Met 42 的侧链可能与 Pc 和 Pp 间脱氧核苷酸 3 个碳原子发生疏水作用。

生化及遗传数据的确认　在阻遏物-操纵基因复合体的详细结构被弄清楚之前，生化和遗传实验已对某些阻遏物氨基酸和操纵基因碱基的重要性做出了预测，并且几乎所有实验都证实了这些预测。

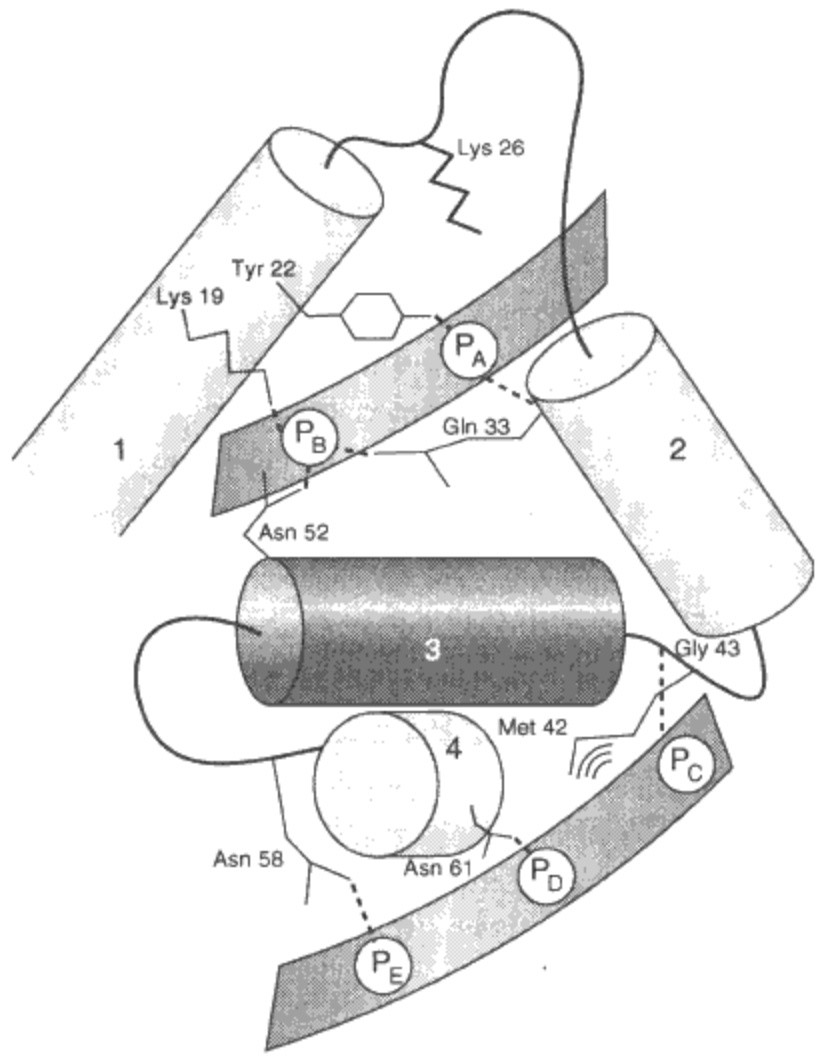


图 9.9　氨基酸/DNA 骨架的相互作用。显示 λ 阻遏物的 α 螺旋 1～4 及其参与形成氢键的磷酸基团（P_A～P_E）。本图与图 9.8 垂直。给出了重要氨基酸的侧链。两条虚线表示肽 NH 基团与磷酸间的氢键。（*Source*：Adapted from Jordan S. R. and C. O. Pabo, Structure of the lambda complex at 2.5 Å resolution: Details of the repressor-operator interactions. *Science* 242：897，1988.）

第一，将操纵基因中的某些磷酸基乙基化可干扰阻遏物的结合。羟基足迹（hydroxyl radical footprinting）实验表明这些磷酸基与阻遏物结合有关。现在已知这些相同的磷酸基（每半位点 5 个）在共结晶中与阻遏物的氨基酸有重要联系。

第二，甲基化保护实验预测 DNA 大沟中某些鸟嘌呤与阻遏物有紧密联系。晶体结构表明这些鸟嘌呤确实与阻遏物的结合有关。大沟鸟嘌呤（G8′）在阻遏物结合时对甲基化较敏感，它在共结晶中有特殊的构象，碱基对 8′水平轴上的扭曲比其他碱基严重，与其他碱基间的距离变宽。这种特殊构象使 8′鸟嘌呤更易受甲基化试剂 DMS 的攻击。在以前的实验中发现过腺嘌呤被甲基化现象，这一结果是合理的，因为腺嘌呤的甲基化发生在位于小沟的 N3

上，而阻遏物与操纵基因的结合不在小沟，所以阻遏物不能保护腺嘌呤免受甲基化。

第三，DNA 测序结果表明在操纵基因 O_R 和 O_L 的所有 12 个半位点中，2 位的 A-T 和 4 位的 G-C 都是保守的，晶体结构揭示它们都参与和阻遏物的重要结合，所以非常保守。

第四，遗传学结果表明，突变某些氨基酸可使阻遏物与操纵基因间的相互作用不再稳定，而改变其他氨基酸却能加强二者的结合。共结晶结构可显示所有这些突变的结果。Lys4 和 Tyr22 突变对结合的破坏最大，我们现在知道它们与操纵基因有很强的结合，Lys4 与 6 位鸟嘌呤（以及 Asn55）、Tyr22 与 P_A 结合。赖氨酸替代 Glu34 可能是正向突变，在晶体结构中没有发现 Glu34 参与任何与操纵基因形成的重要键，但该位点的赖氨酸可旋转于 P_A 前形成盐桥，增强蛋白质与 DNA 的结合。这个盐桥包含赖氨酸 ε 基团的正电荷和磷酸基团的负电荷。

小结 λ阻遏物片段与操纵基因片段的共结晶结构揭示了蛋白质与 DNA 相互作用的细节情况，出现在 DNA 大沟的氨基酸与 DNA 碱基和 DNA 骨架形成氢键是重要的作用方式。通过两个氨基酸与两个或更多 DNA 位点结合所形成的氢键网可以稳定一些氢键。基于共结晶结构推出的结构与生化和遗传实验结果完全一致。

434 噬菌体阻遏物-操纵基因相互作用的高分辨率分析

Harrison、Ptashne 小组利用 X 射线晶体学方法进行了 434 噬菌体阻遏物-操纵基因相互作用的详细分析。与 λ 共结晶一样，他们采用的晶体不是全长阻遏物和操纵基因，而是包含作用位点的阻遏物和操纵基因的部分片段。以阻遏物蛋白的前 69 个氨基酸多肽替代阻遏物，其中包含螺旋-转角-螺旋的 DNA 结合基序。操纵基因是人工合成的含有阻遏物结合位点的 14bp DNA 片段（图 9.10）。理论上这两个片段可以像完整分子那样结合，并且复合体相对容易结晶。这项工作提供了早期 X 射线晶体在完整阻遏物-操纵基因研究中得不到的重要信息。

与碱基对的作用 图 9.10 总结了位于 434 阻遏物识别螺旋（α3）的 Gln28、Gln29、Gln33 侧链与碱基的相互作用。从图 9.10 的底端开始，注意 Gln28 的 Oε、Nε 与 1 位腺嘌呤的 N6、N7 间有两个可能的氢键（虚线）。可见 Gln29 的 Oε 与 Gln29 在蛋白质骨架上的 NH 基团间的氢键，使其 Nε 直接指向操纵基因 2 位鸟嘌呤上的 O6，形成氨基酸与碱基间的氢键。还应注意，Gln29 的 Cβ 和 Cγ 与 3 位碱基对胸腺嘧啶的 5-甲基之间的范德华力（同心弓形）。

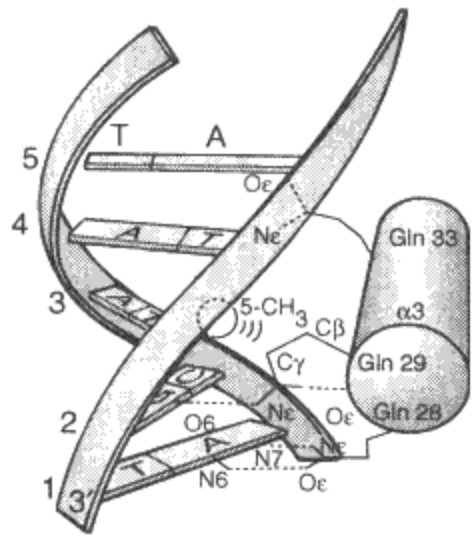


图 9.10 识别螺旋氨基酸侧链与 434 操纵基因半位点相互作用的详细模型。氢键用虚线表示。Gln29 侧链及与 3 位腺嘌呤配对的胸腺嘧啶 5-甲基之间的范德华力相互作用以同心圆表示。（*Source*：Adapted from Anderson，J. E.，M. Ptashne，and S. C. Harrison，Structure of the repressor-operator complex of bacteriophage 434. *Nature* 326：850，1987.）

这种范德华力可初步解释如下：尽管所有参与基团都是非极性的，但在任意给定瞬间由于其电子云的随机波动使基团产生瞬时极性，导致其相邻基团产生相反的极性，结果相邻两基团间产生互相吸引。

小结 434 噬菌体阻遏物片段与操纵基因片段复合体的 X 射线晶体结构表明在氨基酸与 DNA 骨架的磷酸基团间有相互作用形成氢键。该晶体模型也表明在蛋白质的识别螺旋中，3 个谷氨酰胺残基与 3 个碱基对间有氢键生成，其中一个谷氨酰胺与碱基由一个潜在的范德华力连接。

DNA 构象的影响 我们已经讨论过，阻遏物与 DNA 骨架的结合需要 DNA 双螺旋稍微弯曲一下。由 Harrison、Ptashne 及同事进行的高分别率晶体学研究显示，在 DNA-蛋白质复合体中 DNA 确实是弯曲的（图 9.11）。但我们仍不知道 DNA 是事先弯曲的还是由于阻遏物结合后诱导的。在这两种情况下，操纵基因的碱基序列都有助于这种弯曲。就是说，有些 DNA 序列在一定的方式下更容易弯曲，434 操纵基因序列最利于这种弯曲以适应与阻遏物的结合。在本章后面我们将详细讨论这一普遍现象。

操纵基因 DNA 构象的另一个显著特征是两个半位点在 7、8 位碱基对间的双螺旋被压缩为 39°，比正常的 36°多了 3°。在图 9.11b 中可以看到这种影响，注意图中部的小沟狭窄，而两边的大沟比正常的要宽，以补偿这段 DNA 的不完全扭曲。同样，该处的碱基序列最适合这种构象。

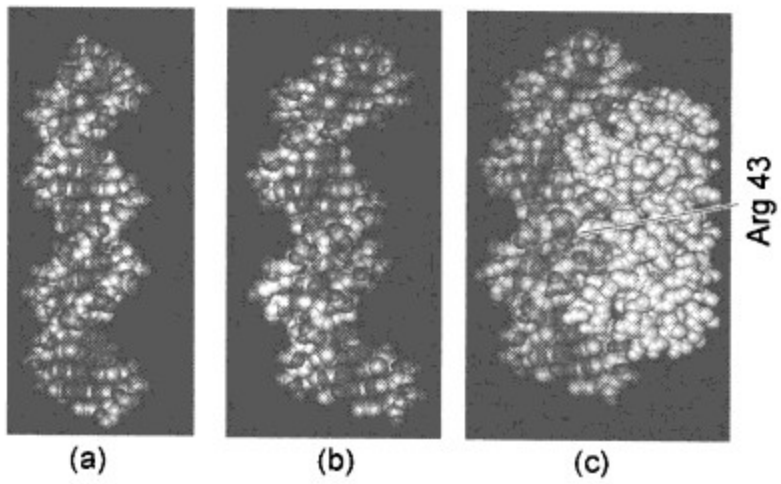


图 9.11　阻遏物-操纵基因复合体中 DNA 弯曲的计算机空间填充模型。（a）标准的 B 型 DNA；（b）阻遏物-操纵基因复合体中含 20mer 片段的操纵基因形状，去除了阻遏物；（c）阻遏物-操纵基因复合体形状，阻遏物为橙色。注意 DNA 是如何适应蛋白质的形状促进两者紧密结合的，可见 Agr43 侧链伸入模型中部 DNA 的小沟。（*Source*：Aggarwal et al.，Recognition of a DNA operator by the repressor of phage 434：A view at high resolution. *Science* 242（11 Nov 1988）f. 3b，f. 3c，p. 902. © AAAS.）

小结　部分 434 噬菌体阻遏物-操纵基因复合体的 X 射线晶体分析表明，DNA 的形状发生了明显变化。它要有一定程度的弯曲以适应碱基/氨基酸相互作用的需要。此外，在两个半位点之间、螺旋中部缠绕得较紧，而其两端则比正常的松。操纵基因的碱基序列有利于这种形状变化。

本模型的遗传学验证　正如我们已经看到的如果阻遏物与操纵基因间的结合很重要，那么突变相关的氨基酸或碱基应当能降低或破坏 DNA-蛋白质结合，或者还可以突变操纵基因使其不能配合阻遏物，然后再做补偿突变恢复其结合。同样，如果操纵基因的特殊形状很重要，那么突变使其不能形成这种形状也应当降低或破坏结合。我们将看到，这些推测都得到了证实。

为了验证 Gln28 和 A1 间相互作用的重要性，Ptashne 小组将 A 突变为 T，期望破坏阻遏物对操纵基因的结合，但该突变又可以被阻遏物 28 位 Gln 到 Ala 的突变所抑制。图 9.10 对此做了可能的解释：Gln28 与 A1 的两个氢键可被 Ala28 与 T1 甲基间的范德华力所取代。这种范德华力的重要性通过不含甲基的尿嘧啶或含甲基的 5-甲基胞嘧啶（5MeC）取代 T1 得以加强。这种 U 取代的操纵基因不能结合含 Ala28 的阻遏物，但可结合 5MeC-取代的操纵基因。因此，甲基基团对于突变体阻遏物结合突变体操纵基因十分重要，这与基于范德华力所做的推测一致。

我们很怀疑碱基对 7、8 间的 DNA 超扭曲（overwinding）对阻遏物-操纵基因相互作用很重要。如果是这样，那么用 G-C 或 C-G 碱基对替换 6～9 位的 A-T 或 T-A 应该能降低阻遏物-操纵基因的结合，因为 G-C 碱基对不能像 A-T 碱基对那样使 DNA 超扭曲。正如所预测的，阻遏物不能很好地结合 6～9 位的 G-C 或 C-G 操纵基因。虽然这不证明超扭曲存在，但与超扭曲的假设一致。

小结　由 X 射线晶体学所预测的 434 噬菌体阻遏物和操纵基因间的相互作用可以被遗传学分析所证实。与结合有关的氨基酸或碱基改变之后，阻遏物-操纵基因的结合受阻。此外，当 DNA 发生突变导致其形状与阻遏物-操纵基因复合体预期的不同时，结合也会被抑制。

9.2　*trp* 阻遏物

trp 阻遏物是另一种利用螺旋-转角-螺旋 DNA 结合基序的蛋白质。回想第 7 章中描述的阻遏物蛋白（没有色氨酸辅阻遏物的蛋白）是非活性的。Paul Sigle 用 *trp* 阻遏物和阻遏物蛋白的 X 射线晶体学研究发现了色氨酸所造成的微小但很重要的差异，这一研究阐明了 *trp* 阻遏物与其操纵基因的作用方式。

色氨酸的作用

这里有一张色氨酸影响阻遏物形状的图示。将色氨酸加入阻遏物蛋白晶体后，晶体破碎了！当色氨酸楔入阻遏物蛋白形成阻遏物时，改变了蛋白质的构型而破坏支撑晶格的结合力。

这产生了一个明显的问题，当色氨酸结合阻遏物蛋白时，什么在移动？为了得到答案，观察图 9.12 所示的阻遏物会有所帮助。该蛋白质实际上是两个相同亚基构成的二聚体，两个亚基形成三域结构（three-domain structure）。每个单体的中心域或“平台”由其 A、B、C、F 螺旋组成，位于 DNA 的右侧，离开 DNA。另外两个结构域在 DNA 的近左侧，由每个单体的 D、E 螺旋组成。

现在回到我们的问题，加入色氨酸后，什么东西发生了移动？很明显，平台保持稳定，而另两个结构域发生倾斜，如图 9.12 所示。每个单体的识别螺旋是 E 螺旋，当色氨酸结合后其位置发生明显移动。在顶端单体中，E 螺旋从下向上稍做移动，进入操纵基因大沟内，在该位置上与 DNA 完美结合（或“阅读”）。

Sigler 称这些 DNA 阅读基序（DNA-reading motif）为阅读头（reading head），把它们比作磁带录音机中的阅读头或计算机磁盘的驱动磁头。计算机的阅读头可处于两种位置，加载读盘或离开磁盘。

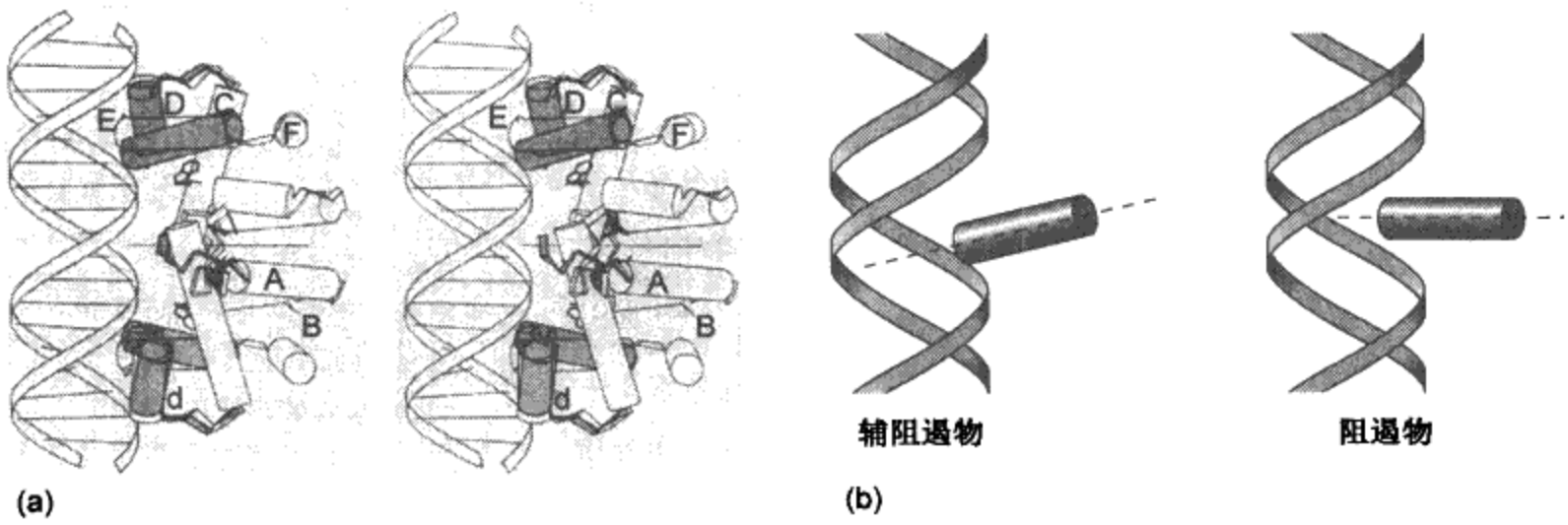


图 9.12　*trp* 阻遏物及阻遏物蛋白与 *trp* 操纵基因匹配的比较。（a）立体图。两个单体的螺旋-转角-螺旋基序在阻遏物（透明的）及阻遏物蛋白（暗的）中应处的位置。阻遏物中色氨酸的位置用黑折线表示。注意阻遏物蛋白识别螺旋（螺旋 E）偏离插入大沟的最佳位置。这两个基本上相同的草图构成空间透视便于立体观察。可用立体镜获得三维效果，或将其举到面前 1～2ft 处，以这个距离盯着，眼睛放松，就可以看见“魔术眼”图片。几秒钟后，两个图在中央汇合成一个三维图。这种立体观察能很好地欣赏识别序列与 DNA 大沟的配合，但是如果只看其中一个图就得不到立体效果。（b）比较阻遏物蛋白（左）和阻遏物（右）识别序列（红色）相对于 DNA 大沟位置的简化图。注意阻遏物识别螺旋指向并直接进入大沟，而阻遏物蛋白识别螺旋的指向偏下。虚线指示识别螺旋的插入角度。

trp 阻遏物以同样方式工作。色氨酸存在时，它将自己插入平台与各阅读头之间，使阅读头处于匹配操纵基因大沟的最适位置，如图 9.12 所示。当色氨酸从阻遏物蛋白上解离后，所留下的空隙允许阅读头离开大沟返回中心平台，离开与操纵基因匹配的位置。

图 9.13a 显示色氨酸在阻遏物上与周围分子的关系。几乎所有可类比的螺旋-转角-螺旋蛋白的疏水口袋都被一个疏水氨基酸（有时是色氨酸）的侧链所占据，如 λ 阻遏物、Cro、CAP。在这些蛋白质中，疏水氨基酸是蛋白链的一部分，不像在 *trp* 阻遏物中是自由氨基酸。Sigler 将 Arg84 和 Arg85 间色氨酸的排列方式比作意大利面肠三明治，其中色氨酸是意大利面肠。将其移走后如图 9.13b 所示，两个精氨酸就像三明治中的两块面包那样贴在一起了。该模型也适用于其他分子，由于 Arg54 位于阻遏物二聚体的中心平台上，而 Arg84 位于阅读头的结合面上，因此在两个精氨酸之间插入色氨酸可将阅读头推离平台而指向操纵基因的大沟（见图 9.12）。

小结　*trp* 阻遏物需要色氨酸迫使阻遏物二聚体的识别螺旋进入适当的位置以便与 *trp* 操纵基因发生相互作用。

9.3　对蛋白质-DNA 相互作用的一般性认识

蛋白质与特定 DNA 片段结合的专一性由什么决定呢？目前所看到的例子提示有两种答案：①碱基与氨基酸的特异性相互作用；②DNA 呈现一定构型的能力，其构型取决于碱基序列。这两种可能性显然互不排斥，适用于很多蛋白质-DNA 的这类相互作用。

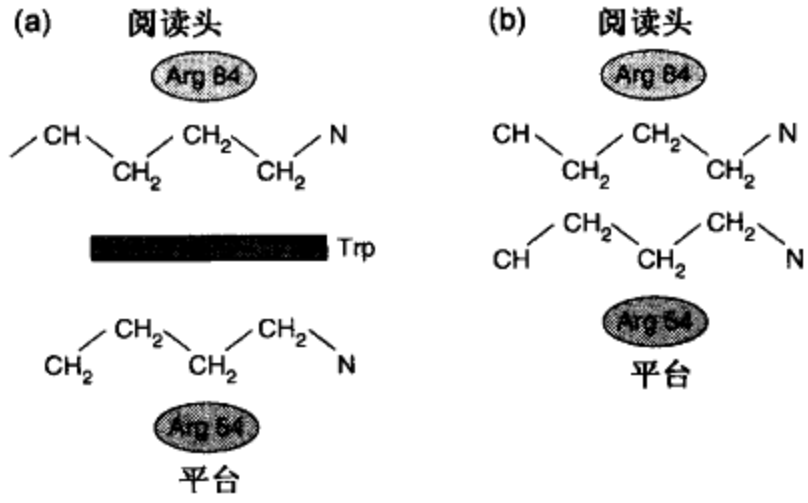


图 9.13　*trp* 阻遏物的色氨酸结合位点。（a）*trp* 阻遏物中色氨酸的邻近分子。注意色氨酸侧链（红色）上方的 Arg84 和下方的 Arg54。（b）阻遏物蛋白没有色氨酸的相同区域。注意精氨酸已经移到一起填补色氨酸缺失留下的空缺。长方形中的氨基酸在阻遏物蛋白的一个亚基上。卵圆形中的氨基酸在另一个亚基上。

4 个不同碱基对的氢键形成能力

我们知道，不同 DNA 结合蛋白依赖于对 DNA 碱基的不同程度的接触。就它们“阅读”碱基序列的程度而言，你可能会问它们读什么？毕

竟碱基对不是敞开的，DNA 结合蛋白必须在碱基配对条件下感知碱基的差异。它们必须通过氢键或范德华力与碱基对产生碱基特异结合（base-specific contact）。我们来查验一下 4 种碱基对的氢键势能。

来看图 9.14a 的 DNA 双螺旋，将 DNA 旋转 90°使其伸出纸面直接指向我们，就可以看到双螺旋轴。现在从这个方向来看一个碱基对，如图 9.14b 所示，大沟在上，小沟在下。DNA 结合蛋白穿过其中一个沟与碱基对发生作用。这样可在沟里“看见”4 个可能的外形，取决于碱基对是 T-A、A-T、C-G 还是 G-C。

图 9.14c 显示来自大小沟的其中两个外形。图 9.14d 总结了蛋白质在两沟与 T-A 和 C-G 相遇的情况。A′表示氢键受体（氧和氮原子），D′表示氢键供体（氢原子）（译者注：氢键受体用 A 表示，供体用 D 表示）。大小沟分别位于水平线之上和之下。竖线的长度表示供体或受体原子从螺旋轴伸向 DNA 沟外的相对距离。可以看到 T-A 和 C-G 碱基对表现出非常不同的图谱，尤其是在大沟里。嘧啶-嘌呤碱基对与嘌呤-嘧啶碱基对的差别在这里更明显。

这些氢键图谱表示了氨基酸与碱基对的直接作用。当然也存在别的可能性，氨基酸还可通过形成氢键或盐键“阅读”DNA 骨架构型。

小结 4 个不同碱基对显示了氨基酸进入 DNA 大、小沟的 4 种不同氢键生成模式。

多聚 DNA 结合蛋白的重要性

Robert Schleif 注意到 DNA 结合蛋白的靶位点通常是对称或重复的，因而可以跟一个以上亚基组成的多聚体蛋白发生作用。大多数 DNA 结合蛋白都是二聚体（有的甚至是四聚体），由此极大地增强了 DNA 与蛋白质的结合，因为蛋白质的两个亚基可协同性地结合。在结合位点只要有一个亚基就自然增加另一个亚基的浓度，这种浓度的迅速增加非常重要，因为细胞内 DNA 结合蛋白的量通常很少。

研究二聚体 DNA 结合蛋白优点的另一个方法是采用熵（entropy）的概念。熵是测量无序状态的一种方式。当你了解到自然界的熵或无序状态随着时间而自然地增加时，可能不会奇怪。想想自己的房间无序时所发生的情况，会对熵的概念有更好的理解。无序状态会随着时间增加，直到你花费能量去收拾。因此将无序变为有序或者降低系统的熵值需消耗能量。

DNA-蛋白质复合体比其相互独立的单个分子更为有序，因此，把它们连在一起降低了熵值。两

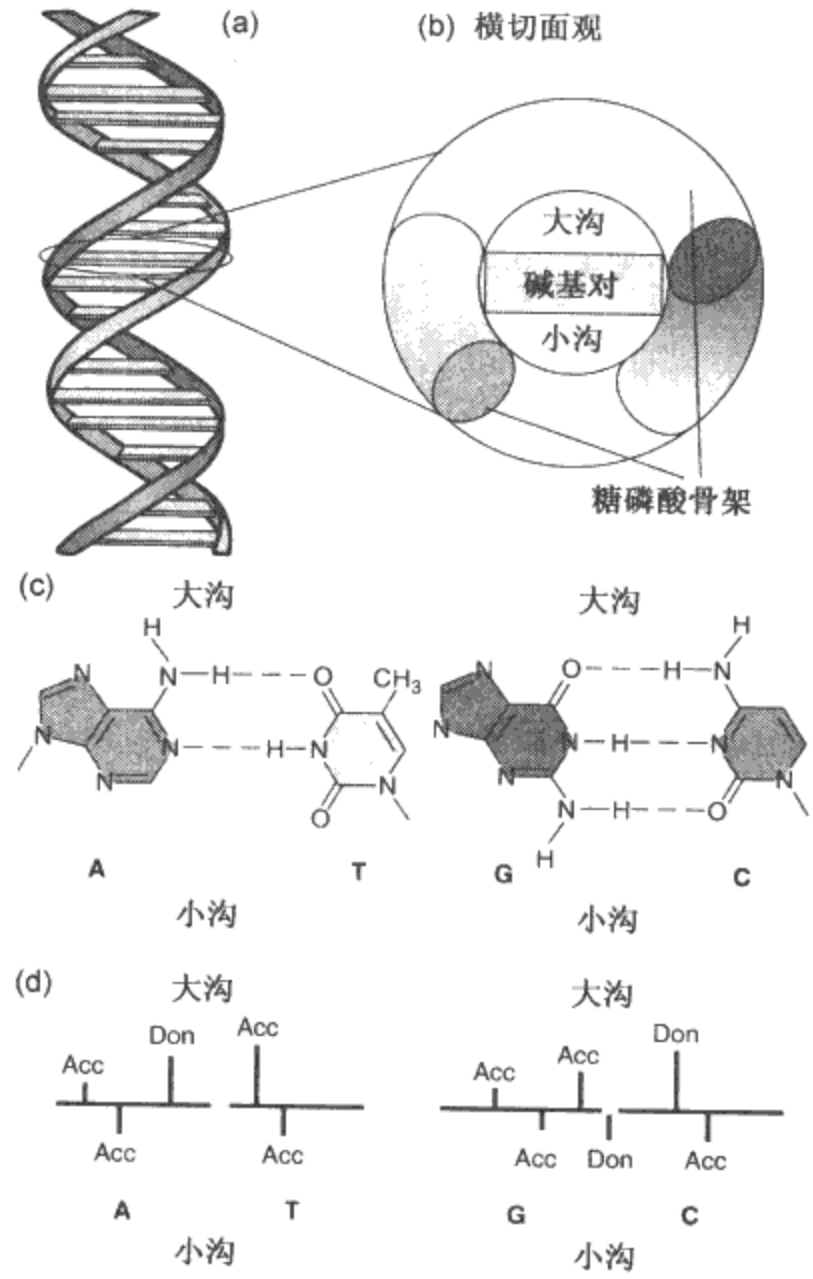


图 9.14 DNA 大小沟内碱基对的外观。（a）标准 B-DNA，两个骨架分别为红色和蓝色，碱基对为黄色。（b）相同 DNA 分子的俯视图。注意与小沟（下）相比大沟（上）开口更大。（c）两个碱基对的结构式。还是大沟在上，小沟在下。（d）大小沟内氢键受体（A）和供体（D）位置的线状图。例如，从左向右 T-A 对的大沟有一个受体（腺嘌呤环的 N-7）一个供体（腺嘌呤的 NH2），再一个受体（胸腺嘧啶的 C═O）。垂直线的交汇点表示这些基团的相对水平位置。竖线的长度表示垂直位置。两个碱基对在大小沟内受体和供体的组合方式不同，因此外部蛋白进入时可识别。将图左右翻转可见 A-T 和 G-C 仍呈不同模式。（*Source*：Adapted from R. Schleif, DNA binding by proteins. *Science* 241：1182-3，1988.）

个相互独立的蛋白亚基结合，可加倍降低熵值。但是如果两个蛋白亚基已经在二聚体中结合，那么调整其中一个与 DNA 的结合，会自动调整另一个的结合，因此熵值变化比单独结合中少得多，所需能量也少。从 DNA-蛋白质复合体的角度看，从 DNA 释放蛋白二聚体的熵值与释放两个独立的结合蛋白

的熵值不等，所以二聚体蛋白和 DNA 结合得更紧密。

小结　多聚 DNA 结合蛋白对 DNA 结合位点的亲和性高于独立结合的多聚单体蛋白。

9.4　DNA 结合蛋白：远程作用

到目前为止，我们主要研究了 DNA 结合蛋白控制其附近事件的发生。例如，*lac* 阻遏物结合到操纵基因上，干扰邻近 DNA 位点的 RNA 聚合酶活性。又例如，λ阻遏物刺激 RNA 聚合酶结合到邻近位点上。然而 DNA 结合蛋白还可以影响远程位点上二者的相互作用。这种现象在真核生物中很普遍，在原核生物中也有几个例子。

Gal 操纵子

1983 年 S. Adhya 与同事报道了一个意外发现，即 *E. coli* 的 *gal* 操纵子（编码半乳糖代谢所需的酶）有两个相隔 97bp 的不同操纵基因。其中一个邻接 *gal* 启动子位于预期的操纵基因的位置，称作 O_E（表示“外部的”操纵基因）。另一个称为 O_I（表示“内部的”操纵基因），位于第一个结构基因 *galE* 之内。O_I 是通过遗传学研究发现的，即在 *galE* 基因而不是 O_E 定位到 O^c 突变体。对这种分开的两个操纵基因功能的一种解释是，它们都与阻遏物结合，而两个阻遏物则通过 DNA **环凸**（looping out）发生相互作用，如图 9.15 所示。我们在第 7 章已看到 *ara* 操纵子就是采用这种作用机制的。

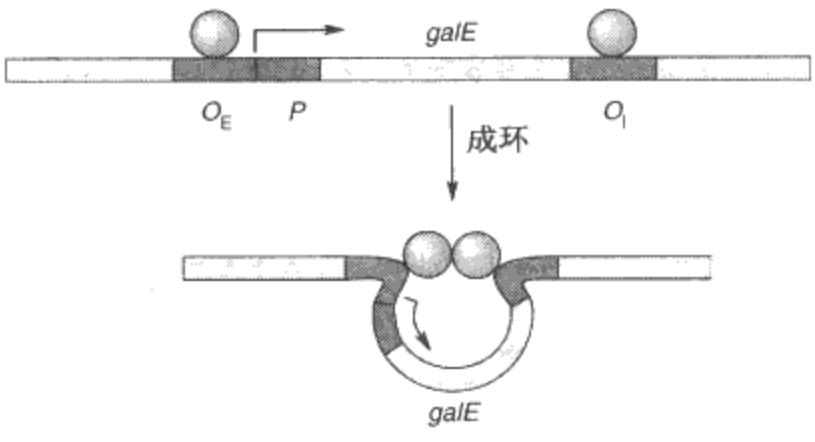


图 9.15　*gal* 操纵子的阻抑作用。 *gal* 操纵子有两个操纵基因（红色）。外部的（O_E）邻接启动子（绿色），内部的（O_I）位于 *galE* 基因（黄色）之内。阻遏物分子（蓝色）结合到两个操纵基因上，通过成环外凸中间的 DNA 发生相互作用。

重复的 λ 操纵基因

刚才对 *gal* 操纵子的简要介绍显示蛋白质在大约 100bp 远的距离上可相互作用，但是没有给出直接证据。Ptashne 及同事用人工体系获得了直接证据。该系统是我们熟悉的 λ 操纵基因-阻遏物组合，但是人为地将相邻的两个操纵基因按不同距离隔开了。正常情况下，当两个操纵基因 O_R1 和 O_R2 相邻时，阻遏物二聚体协同性地与它们结合。问题是当两个操纵基因分开后，阻遏物二聚体还能协同地与它们结合吗？答案是肯定的，只要两个操纵基因位于 DNA 双螺旋的同一侧就可以。这一发现支持了阻遏物与 *gal* 操纵基因的结合由 DNA 环介导的假设。

Ptashne 小组用 DNase 足迹和电子显微镜观察证明了隔开的 λ 启动子上阻遏物的协同结合。如果对两个独自结合相隔较远 DNA 位点的蛋白质进行 DNase 足迹实验，可以看到两个独立的足迹。如果对通过 DNA 成环协同结合相隔较远 DNA 位点的两个蛋白质进行足迹实验，同样可以看到两个独立的足迹，但还能看到另外一个有趣现象，即对 DNase 先是不敏感然后是超敏感的交替重复图案。对这一现象可用图 9.16 解释，当 DNA 成环外凸时，内侧的碱基对被压紧，相对地可免受 DNase 的破坏，而环外侧的碱基对则比正常情况要伸展，对 DNA 酶变得更敏感。随着 DNA 双螺旋延伸，这种模式就交替重复出现。

利用这种方法分析协同性，Ptashne 他们对结合有 DNA 的阻遏物做了 DNA 酶足迹实验，实验中的两个操纵基因被整数或非整数的双螺旋数所分开。图 9.17a 显示两个操纵基因被隔开 63bp（几乎刚好是 6 个螺旋）时阻遏物协同结合的例子，可以看到在两个结合位点间对 DNA 酶敏感性高低变化交替重复。图 9.17b 是非协同结合的结果，两个操纵基因被 58bp（刚好 5.5 个螺旋）所分开，没有看到对 DNase 敏感性的交替变化。

他们又对阻遏物-操纵基因复合体进行电子显微镜观察，看阻遏物中结合的 DNA 是否真的成环外凸了，其中的操纵基因被整数或非整数螺旋所隔开。图 9.18 的结果显示 DNA 弯曲很严重，成环外凸正在发生。但在非整数螺旋所隔开的操纵基因中没有看到 DNA 弯曲，这些 DNA 很难成环外凸。这些实验清楚地证明当两个 DNA 位点被整数螺旋隔开时，通过位点间 DNA 的成环外凸而使蛋白质协同结合。

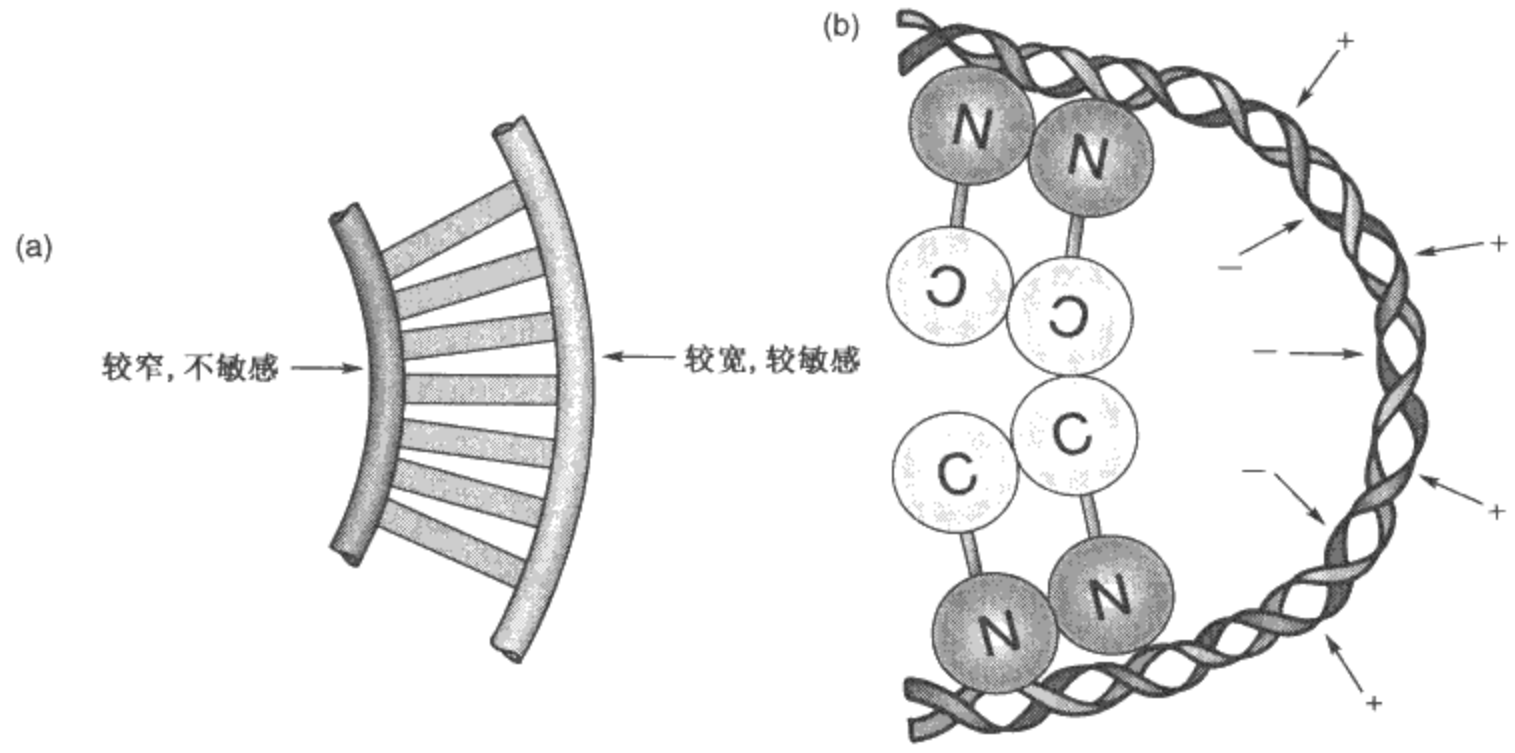


图 9.16　DNA 成环对 DNase 敏感性的影响。（a）简化的示意图。双螺旋简化为铁轨形式，骨架为红色和蓝色，碱基对为橙色。当 DNA 弯曲时，内侧链被压紧，限制了 DNase 的进入，而外侧链被拉开，易受 DNase 攻击。（b）在真实双螺旋中，两条链在内外侧交替出现。DNA 结合蛋白二聚体（本例是 λ 阻遏物）在两个分开的位点上相互作用，使中间的 DNA 成环外凸，使环外侧的 DNA 伸展，易受 DNaseI 攻击（＋标出），而环内侧 DNA 被压缩，阻碍 DNaseI 攻击（－标出）。结果出现了成环区对 DNaseI 敏感性较高和较低的交替方式。这里只考虑了一条链（红色），对另一条链也同样适用。（*Source*：（*b*）Adapted from Hochschild A. and M. Ptashne，Cooperative binding of lambda repressors to sites separated by integral turns of the DNA helix. *Cell* 44：685，1986.）

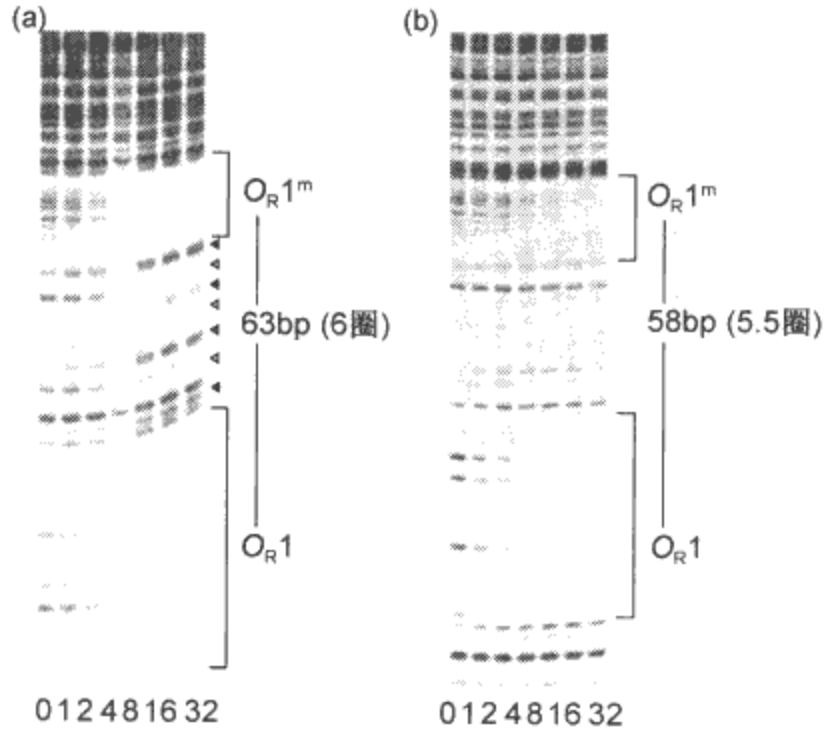


图 9.17　双操纵基因位点的 DNase 足迹。（a）协同结合。两个操纵基因几乎刚好相距 6 个双螺旋（63bp），加入的阻遏物的浓度增加时，DNaseI 在两个足迹间强弱交替切割。实心箭头表示强的，空心箭头表示弱的。由此提示阻遏物结合的两个操纵基因间 DNA 成环。（b）非协同结合。两个操纵基因被非整数双螺旋（58bp 或 5.5 圈）所隔开。阻遏物结合时，对 DNase 的敏感性未表现强弱交替，因此两个阻遏物分别独立地结合在操纵基因上，没有 DNA 成环。每个泳道底部的数字是加入的阻遏物单体的浓度，1 是 13.5nmol/L，2 是 27nmol/L 等。（*Source*：Adapted from Hochschild，A. and M. Ptashne，Cooperative binding of lambda repressors to sites separated by integral turns of the DNA helix. *Cell* 44（14 Mar 1986）f. 3a&4，p. 683.）

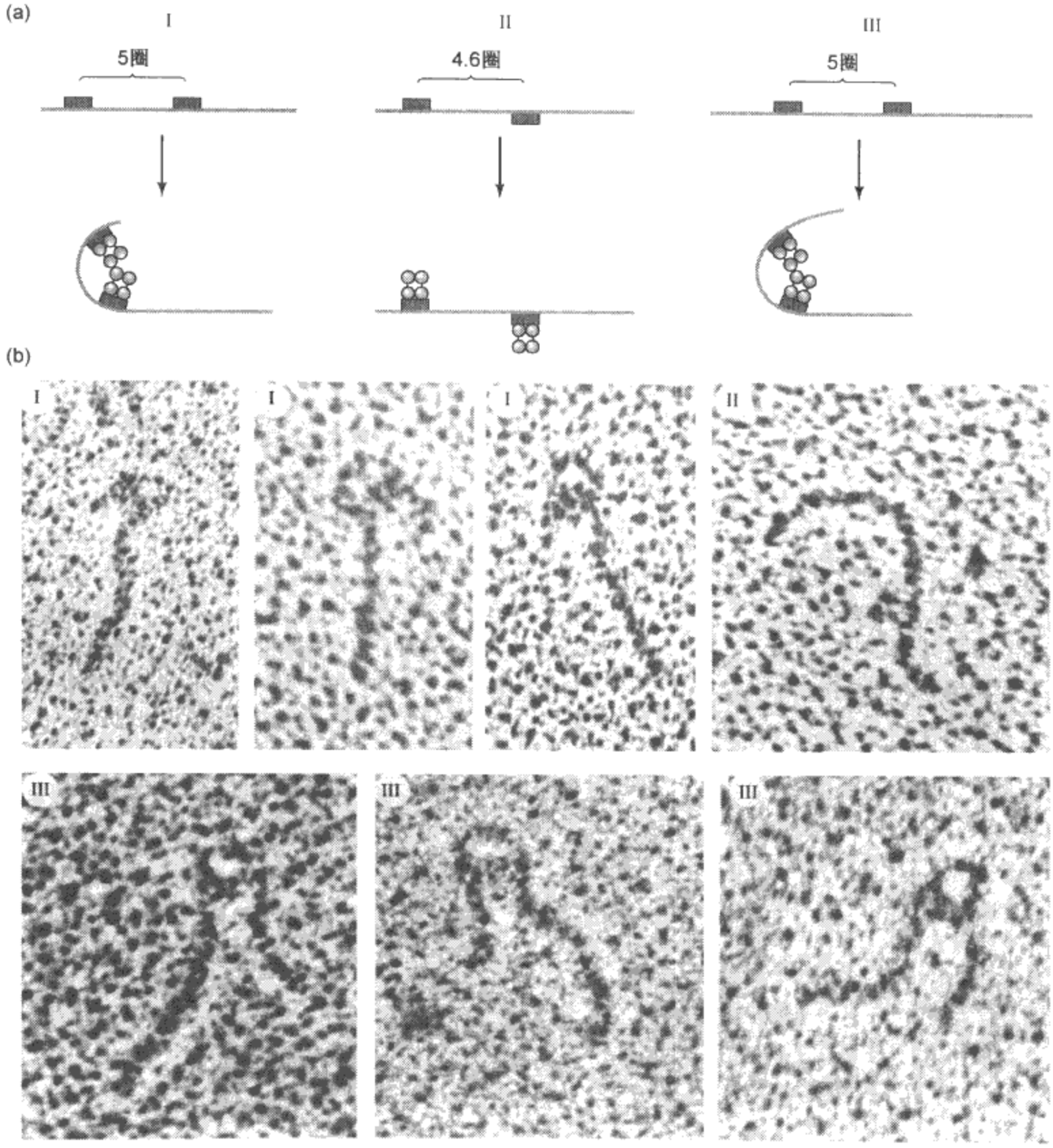


图 9.18　结合于双操纵基因的入阻遏物的电镜照片。（a）三个 DNA 分子的双操纵基因排列。图 I，两个操纵基因在 DNA 一端相距 5 个螺旋转角；图 II，在 DNA 一端相距 4.6 个螺旋转角；图 III，在 DNA 中间相距 5 个螺旋转角；箭头指示每种情况下由于阻遏物与两个操纵基因的协同性结合所应产生的 DNA 环的形状。在图 II 不应有环产生，因为间隔的螺旋数是非整数，阻遏物结合在 DNA 双链的两个相反面。（b）蛋白质-DNA 复合物的电镜照片。所用 DNA 的类型在每张图的左上角标出。复合物确实具有图（a）所示形状。（*Source*：（a）Griffith et al.，DNA loops induced by cooperative binding of lambda repressor. *Nature* 322（21 Aug 1986）f. 2，p. 751. © Macmillan Magazines Ltd.）

小结　当两个 λ 操纵基因被整数螺旋隔开时，其间的 DNA 成环外凸，使蛋白质协同结合。当操纵基因被非整数螺旋隔开时，蛋白质只能结合到 DNA 双螺旋的对面，所以不能发生协同结合。

增强子

增强子是非启动子 DNA 元件，可与蛋白质因子结合激活转录。根据定义，增强子可以远程发挥作用。1981 年以来，已在真核生物中确认这一元件，我们将在第 12 章中详述。最近，增强子在原核

生物中也被发现了。1989 年 Popham 及同事描述了一种增强子，它能帮助 *E. coli* 中辅助性 σ 因子（σ^{54}）所识别的基因的转录。我们在第 8 章中提过这个因子，它是个 σ 因子，也称为 σ^{N}，在氮源缺乏时发挥作用，可以从另一个启动子处转录 *glnA* 基因。

σ^{54} 因子是缺陷型的。DNase 足迹实验证明，它可引起 $E\sigma^{54}$ 全酶稳定结合到 *glnA* 启动子上，但不能像正常 σ 因子那样发挥形成开放启动子复合体的重要功能。Popham 等利用**发夹阻抗**（hairpin resistance）和 DNA 甲基化两种方法分析了它的功能。当聚合酶形成开放启动子复合体时，与 DNA 结合得非常紧密。加入发夹结构作为 DNA 的竞争者不能阻止聚合酶。而聚合酶形成的闭合启动子复合体是一种相对松散的结合并以很高的速度解离，因此，加入发夹竞争剂可以抑制聚合酶结合 DNA。此外，当聚合酶形成开放启动子复合体时，会使变性 DNA 中的胞嘧啶暴露在 DMS 中发生甲基化。而闭合复合体中无 DNA 解链，所以不会发生甲基化。

通过发夹敏感性和甲基化抗性这些关键实验发现，σ^{54} 不能形成开放复合体，而是由另一个蛋白质 NtrC（*ntr*C 基因的产物）结合到增强子上帮助 σ^{54} 形成开放复合体。DNA 解链所需能量来自 ATP 水解，而 ATP 水解由 NtrC 上的 ATP 酶结构域行使。这种 ATP 依赖型 σ 因子在原核生物中不多见。通常，非增强子依赖型启动子的 σ 因子可使 DNA 在没有 ATP 的情况下解链。

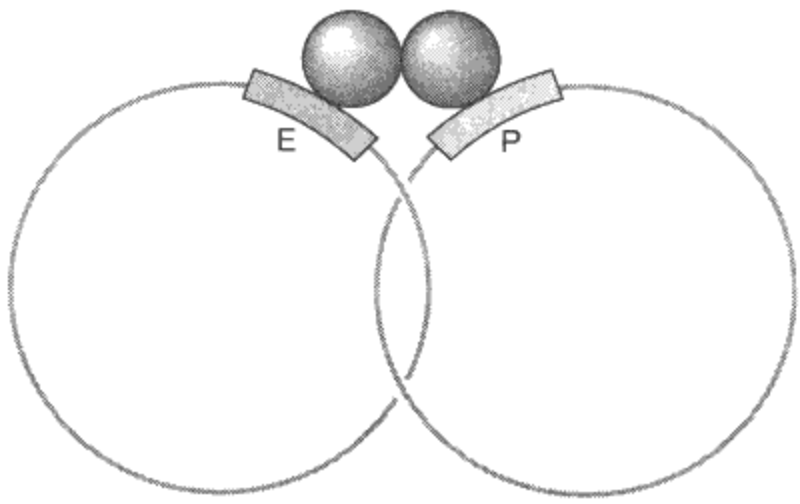


图 9.19　两个处于联结 DNA 分子上的分离位点的相互作用。增强子（E）和启动子（P）分别位于两个拓扑性连体环 DNA 分子上。因此，即使两个环有差异，增强子和启动子永远不会离开太远，从而使结合在上面的蛋白质（红色和绿色）发生相互作用。

那么增强子与启动子如何作用呢？实验结果有力地提示有 DNA 成环参与其中。一条线索是，增强子必须离开启动子至少 70bp 才能发挥作用，这样增强子与启动子间有足够的空间使 DNA 成环。此外，即使增强子与启动子不在同一个 DNA 上也能发挥作用，前提是两者连接成图 9.19 所示的连体环。这样仍能使增强子与启动子像在成环中那样发生相互作用，从而排除了要求两个元件必须在同一 DNA 上的机制（如超螺旋程度的改变或蛋白质沿 DNA 滑动）。在第 12 章我们将详细讨论这个现象。最后，也许最有说服力，我们可以实际观察预测的 DNA 环，它位于 NtrC 结合的增强子和 σ^{54} 全酶结合的启动子之间。图 9.20 是 Sydney Kustu 和 Harrison Echols 及同事的电子显微镜实验结果。所用 DNA 克隆含有增强子-*glnA* 区域，他们在启动子与增强子间插入 350bp DNA，以便成环更容易观察。在多数电子显微镜图片上，聚合酶全酶比 NtrC 着色要深，因此在环的基部可区分两个蛋白质，就像我们预测两种蛋白质是否通过其间 DNA 成环而相互作用一样。

T4 噬菌体提供了一个非正常的可移动增强子的例子，它不是根据一组碱基序列来定义的。T4 噬菌体晚期基因的转录依赖于 DNA 复制。只有噬菌体 DNA 开始复制后，晚期基因才能转录。晚期转录与 DNA 复制间关联的原因之一是噬菌体晚期 σ 因子（σ^{55}）与 *E. coli* 的 σ^{54} 一样是缺陷型的，没有增强子它就不能行使功能。但是晚期 T4 增强子不像 NtrC 结合位点，不是一段固定 DNA 序列，而是一个 DAN 复制叉。增强子结合蛋白由噬菌体基因 44、45 和 62 编码，是噬菌体 DNA 复制机器的一部分，增强子结合蛋白随复制叉移动，与移动的增强子保持接触。

可以在体外用一个简单的 DNA 切口模拟复制叉，但这个切口的极性十分重要，只有位于非模板链时才能起增强子作用。由此提示 T4 晚期增强子可能不是通过 DNA 成环发挥作用，因为成环与极性无关。此外，不像 *glnA* 等典型的增强子，T4 晚期增强子必须与它所控制的启动子在同一个 DNA 分子上。它也不能作为连体环的一部分发挥反式（*in trans*）作用，这有悖于成环机制。

小结　*E. coli* 的 *glnA* 基因是依赖增强子进行转录的原核生物的例子。增强子结合 NtrC 蛋白，而 NtrC 蛋白再结合至少 70bp 之外的启动子上的聚合酶，通过 NtrC 的 ATP 水解提供能量，形成开放启动子复合体起始转录。两个蛋白质通过使中间的 DNA 成环外凸而发生相互作用。T4 晚期增强子是 DNA 复制复合体的一部分，可移动。由于它必须与所控制的启动子在同一 DNA 分子上，因此可能不是通过 DNA 成环发挥作用。

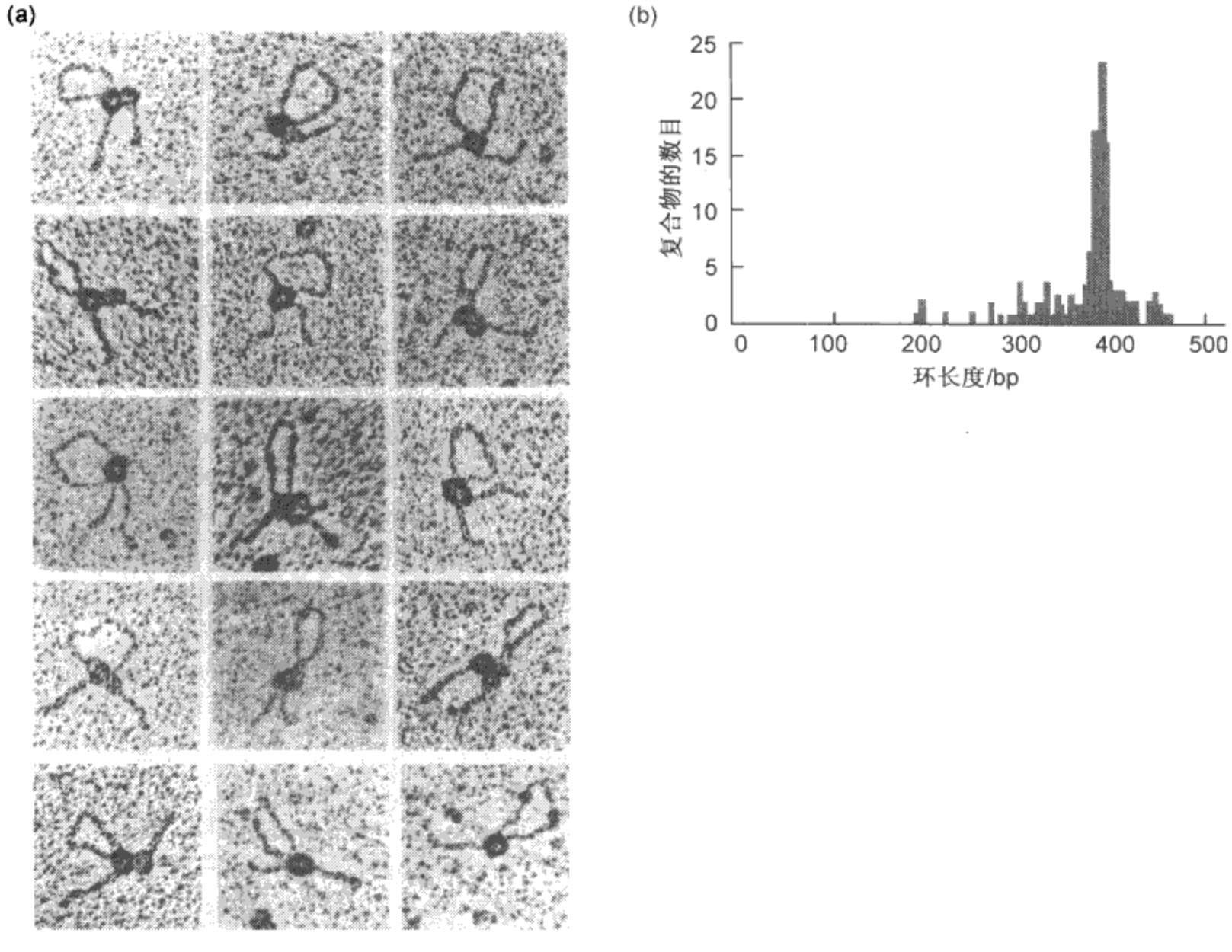


图 9.20　*glnA* 启动子-增强子区的成环作用。(a) 电镜照片。Kustu、Echols 小组在启动子与增强子间插入 350bp DNA 片段使两者分开，然后让 NtrC 结合到增强子上、RNA 聚合酶结合到启动子上。当两个蛋白质相互作用时，中间的 DNA 成环外凸，如电镜照片所示。(b) 环长频数图。测量 (a) 中环的长度，将环长与环数作图，峰值为 390bp，正是启动子与增强子之间的距离。表明环不是 DNA 链交叉形成的假象，而是由两个远程蛋白质之间相互作用所形成的真实环。(*Source*: (a) Su, W., S. Porter, S. Kustu, and H. Echols, DNA-looping and enhancer activity: Association between DNA-bound NtrC activator and RNA polymerase at the bacterial *glnA* promoter. *Proceedings of the National Academy of Sciences USA* 87 (July 1990) f. 4, p. 5507.)

总结

类 λ 噬菌体阻遏物都具有识别螺旋可从侧面契合到操纵基因 DNA 大沟里。识别螺旋靠近 DNA 侧的某些氨基酸可特异地结合操纵基因的碱基，这种结合决定了蛋白质-DNA 相互作用的特异性。改变这些氨基酸可以改变阻遏物的特异性。λ 阻遏物和 Cro 蛋白对相同操纵基因具有相同的亲和性，但它们对 O_R1 或 O_R3 的亲和力具有细微差别，这种差别由识别螺旋中不同氨基酸与两个操纵基因中不同碱基对的相互作用决定。

λ 阻遏物片段与操纵基因片段的共晶体结构揭示了蛋白质与 DNA 相互作用的细节情况。最重要的结合发生在 DNA 大沟，是识别螺旋的氨基酸与 DNA 碱基和 DNA 骨架形成氢键。由两个氨基酸与两个或更多 DNA 位点结合所形成的氢键网络可以稳定一些氢键。通过共晶体结构推测的结构与生化和遗传实验结果几乎完全一致。

434 噬菌体阻遏物片段/操纵基因片段复合体的 X 射线晶体显示了识别螺旋的氨基酸与操纵基因（译者注：原文为阻遏物）碱基对之间的可能氢键结合，也揭示了识别螺旋的氨基酸与操纵基因碱基对之间的潜在范德华力结合。复合体中 DNA 形状发生了明显变化，它适当弯曲以适应碱基/氨基酸的结合。此外，在两个半位点间的螺旋中部缠绕格外紧，而外侧比正常的要松。操纵基因的碱基序列有利于其与正常 DNA 构型的差异。

trp 阻遏物需要色氨酸迫使阻遏物二聚体的识

别螺旋进入适当的位置与 *trp* 操纵基因相互作用。

DNA 结合蛋白可以结合 DNA 大沟或小沟（或两个同时）。4 个不同碱基对呈现 4 种不同氢键模式，与指向 DNA 大小沟的氨基酸结合。

多聚 DNA 结合蛋白对 DNA 结合位点的亲和性高于独立结合的多聚单体蛋白。多聚蛋白的优势是可以协同性地与 DNA 结合。

当两个 λ 操纵基因被整数螺旋隔开时，其间的 DNA 成环外凸，允许协同结合。当操纵基因被非整数螺旋隔开时，蛋白质只能结合到 DNA 双螺旋的对面，所以不能发生协同结合。

E. coli 的 *glnA* 基因是依赖于增强子转录的原核生物的例子。增强子与 NtrC 蛋白结合，而 NtrC 蛋白再与至少 70bp 之外的启动子上的聚合酶结合，通过 NtrC 的 ATP 水解提供能量，形成开放启动子复合体起始转录。两个蛋白质通过成环中间外凸的 DNA 而相互作用。T4 晚期增强子是 DNA 复制复合体的一部分，可移动。由于它必须与所控制的启动子在同一个 DNA 分子上，因此可能不是通过 DNA 成环发挥作用。

复习题

1. 图示与 DNA 双螺旋相互作用的螺旋-转角-螺旋结构域。

2. 描述实验并给出结果，显示在类 λ 噬菌体阻遏物与其操纵基因结合中起重要作用的氨基酸。给出两种方法分析阻遏物与操纵基因的结合。

3. 总体来讲，什么因素决定 λ 阻遏物和 Cro 对 3 个操纵基因位点的偏爱性？

4. 谷氨酰胺和精氨酸侧链与 DNA 之间倾向于产生何种键？

5. 氨基酸的亚甲基和甲基基团与 DNA 之间倾向于产生何种键？

6. 在蛋白质-DNA 相互作用中，氢键网络是什么意思？

7. 图示“阅读头”模型，显示相对于 *trp* 操纵基因，*trp* 阻遏物及阻遏物蛋白识别螺旋的位置不同。

8. 图示“意大利三明治”模型，解释将色氨酸加入 *trp* 阻遏物蛋白后如何引起蛋白质构象的转换？

9. 用一句话对比 λ 和 *trp* 阻遏物相对于各自操纵基因的方向。

10. 解释为什么寡聚蛋白（二聚体或四聚体）与 DNA 的结合比单体蛋白质更容易？

11. 图解当两个蛋白质协同性地结合到 DNA 的两个分开位点上时，对 DNase 抗性和敏感性的交替模式。

12. 描述实验并给出结果，证明 λ 阻遏物二聚体与两个被整数双螺旋隔开的操纵基因的结合是协同性的，而与两个被非整数双螺旋隔开的操纵基因的结合是非协同性的。

13. 描述并给出电镜实验结果，证明与上题中的问题相同的结果。

14. σ^{54} 的缺陷型以何种方式表现？

15. 哪些物质能补偿 σ^{54} 缺失的功能？

16. 叙述证明 DNA 成环参与 *E. coli* 的 *glnA* 基因位点增强作用的实验和结果。

17. T4 噬菌体 σ^{55} 的增强子与 *E. coli* σ^{54} 的增强子有何不同？

分析题

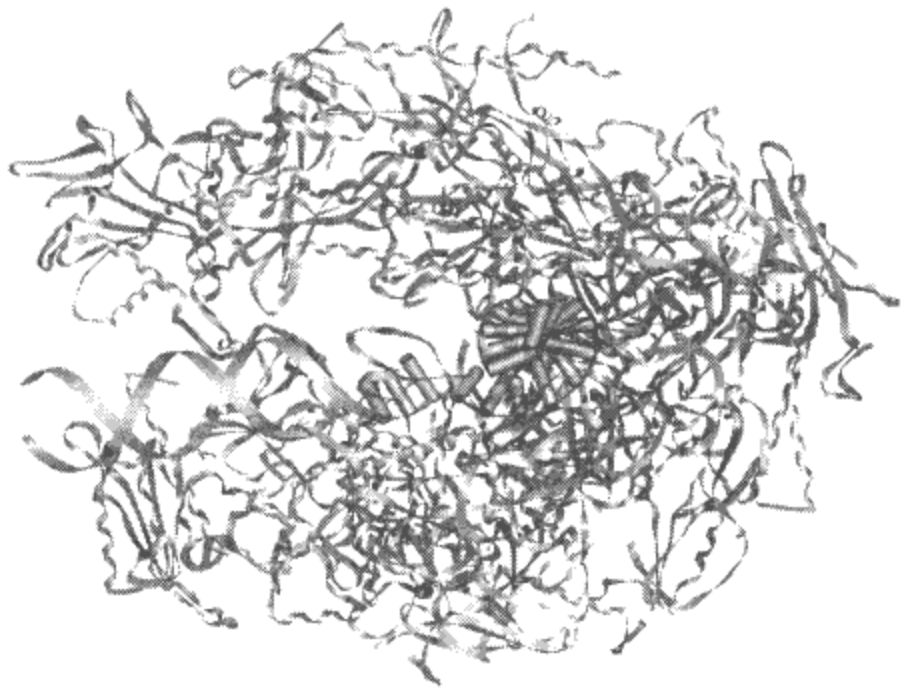
1. DNA 结合蛋白的一个精氨酸与 DNA 的胞嘧啶形成一个重要的氢键。将这个精氨酸（译者注：原文为谷氨酰胺）改变为丙氨酸可阻止这个氢键的形成，从而阻断 DNA-蛋白质之间的相互作用。将胞嘧啶变为胸腺嘧啶可恢复对突变蛋白的结合。给出一个可接受的假说解释这种发现。

2. 你有以下假说：要很好地与 DNA 结合蛋白结合，DNA 靶位点必须螺旋得松一些，加宽碱基对 4 和 5 之间的小沟。设计一个实验验证这个假设。

3. 图示 A-T 碱基对结构。以此结构为基础，画出线状图，示意氢键供体和受体基团在大小沟中的相对位置。碱基对的水平轴用两个水平线段表示，氢键供体和受体的相对水平位置用竖线表示。氢键供体和受体的相对垂直位置用竖线长度表示。当蛋白质与该碱基对相互作用时，该图具有怎样的相关性？

翻译　马　纪　校译　张富春

第 10 章　真核生物的 RNA 聚合酶及其启动子



活性位点结合有 RNA-DNA 杂合分子的酵母 pol IIΔ4/7 蛋白的计算机生成模型。
（*Source*: *David A. Bushnell*, *Kenneth D. Westover*, *and Roger D. Korberg.*）

在第 6 章，我们知道细菌只有一种 RNA 聚合酶，负责转录三类 RNA：mRNA、rRNA 和 tRNA。事实上，该聚合酶可通过切换 σ 因子而满足环境变化的需要，但核心酶的本质是相同的，这完全不同于真核生物的情形。在本章我们将了解到，在真核细胞的细胞核中存在三类截然不同的 RNA 聚合酶，每种 RNA 聚合酶均负责转录一套独立的基因，并识别不同的启动子。

10.1　真核生物 RNA 聚合酶的多种形式

早期研究认为，在真核生物细胞核内至少有两类 RNA 聚合酶在发挥作用：一类负责转录核糖体 RNA 基因（在脊椎动物中，这些基因编码 28S、18S 和 5.8S rRNA）；另一类或多类 RNA 聚合酶负责转录细胞核内其他的基因。

首先，核糖体基因在以下几方面与细胞核内其他基因存在着差异。①核糖体基因与其他核基因的碱基组成不同，如小鼠 rRNA 基因的 GC 含量为 60%，而其他 DNA 序列的 GC 含量仅为 40%。②显著重复，其重复程度因生物种类而异，每个细胞 rRNA 基因的拷贝数从几百到 2 万以上不等。③核糖体基因与其他核基因在细胞核内的定位不同，rRNA 基因存在于核仁（nucleolus）中。这些特征都暗示，真核生物细胞核内至少存在两类功能性 RNA 聚合酶，一类存在于核仁中，负责 rRNA 的合成；另一类存在于核质（nucleoplasm）（核仁之外的其他细胞核部分）中，负责其他 RNA 的合成。

三类细胞核聚合酶的分离

1969 年，Robert Roeder 和 William Rutter 指出，真核生物拥有三类而不是两类不同的 RNA 聚合酶。而且，这三类 RNA 聚合酶在细胞内具有各自截然不同的功能。研究者用 DEAE-Sephadex 离子交换层析方法（第 5 章）从真核生物细胞中分离出了这 3 类不同的 RNA 聚合酶。

RNA 聚合酶 I 定位于核仁中，催化合成 rRNA；RNA 聚合酶 II 和 RNA 聚合酶 III 定位于核质中，催化合成其他的 RNA。

根据从离子交换柱上洗脱的次序，研究者将聚合酶活性的 3 个峰值命名为 **RNA 聚合酶 I、RNA 聚合酶 II 和 RNA 聚合酶 III**（图 10.1）。这三类 RNA 聚合酶除在 DEAE-Sephadex 离子交换层析上表现出不同行为外，还有其他不同的性质。例如，这三类 RNA 聚合酶对不同的离子强度及二价金属离子具有不同反应特性，更为重要的是，它们在基因的转录事件中各自发挥特定的功能，即每一种 RNA 聚合酶转录合成特定的一类 RNA。

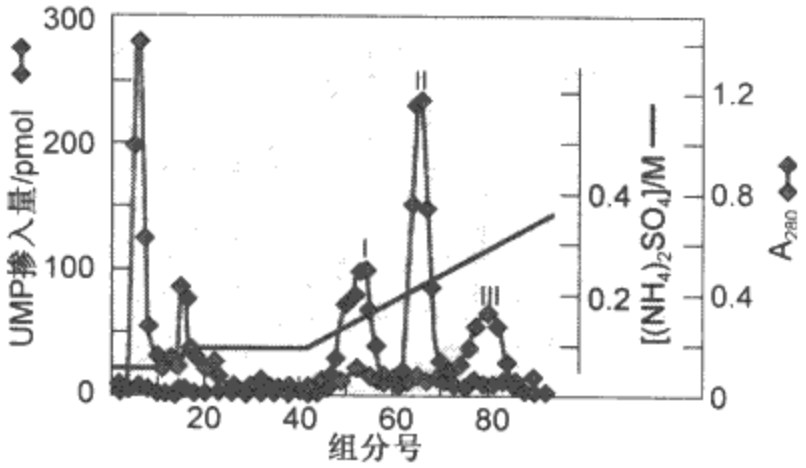


图 10.1　真核生物 RNA 聚合酶的分离。Roeder 和 Rutter 对海胆胚提取物进行 DEAE-Sephadex 离子交换层析。绿色：在 A_{280} 波长测得的蛋白质吸光值；红色：通过检测 RNA 分子中标记 UMP 掺入量而测得的 RNA 聚合酶活性；蓝色：硫酸铵浓度。（*Source*：Adapted from Roeder，R. G.，and W. J. Rutter，Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature* 224：235，1969.）

Roeder 和 Rutter 进一步研究了分离纯化的核仁和核质，以查看这些亚细胞核区室（subnuclear compartment）是否富集有相应的聚合酶。图 10.2 显示，聚合酶 I 确实主要定位在核仁中，聚合酶 II 和聚合酶 III 则存在于核质中。说明聚合酶 I 极有可能就是 rRNA 合成酶，聚合酶 II 和聚合酶 III 是其他 RNA 的合成酶。

小结　真核生物细胞核内存在三类 RNA 聚合酶，可通过离子交换层析的方法进行分离。RNA 聚合酶 I 定位于核仁中，RNA 聚合酶 II 和 RNA 聚合酶 III 定位于核质中。RNA 聚合酶 I 在核仁中的定位表明其负责 rRNA 基因的转录。

三类 RNA 聚合酶的作用

我们如何分辨这三类 RNA 聚合酶在转录中具有不同的作用？有关聚合酶作用的最终证据是通过体外实验获得的，即研究纯化 RNA 聚合酶在体外条件下只能转录哪些特定基因，而不能转录其他基因。结果表明，三类 RNA 聚合酶具有如下特性（表 10.1）。RNA 聚合酶 I 催化合成大的 rRNA 前体，在哺乳动物中，这个前体的沉降系数为 45S，可被加工为成熟的 5.8S、18S 和 28S rRNA。RNA 聚合酶 II 催化合成两类 RNA，即**核内不均一 RNA**（heterogeneous nuclear RNA，hnRNA）以及大多数的**核内小分子 RNA**（small nuclear RNA，snRNA）。对于 hnRNA 的界定目前还不明确。在第 14 章我们将会了解到，绝大多数 hnRNA 是 mRNA 的前体，而 snRNA 则参与了由 hnRNA 到 mRNA 的成熟过

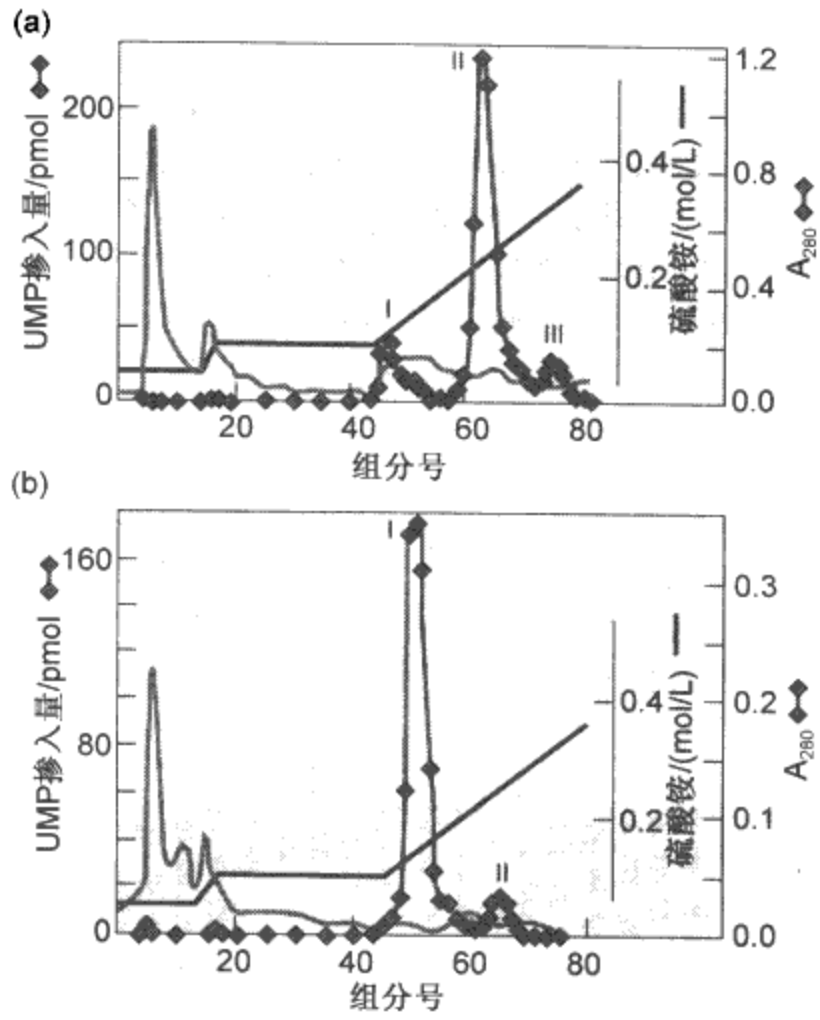


图 10.2　大鼠肝脏细胞内三类 RNA 聚合酶的细胞定位。Roeder 和 Rutter 将小鼠肝脏细胞的核质馏分（a）或核仁馏分（b）进行 DEAE-Sephadex 离子交换层析，方法及曲线颜色均同图 10.1。（*Source*：Adapted from Roeder，R. G.，and W. J. Rutter，Specific nucleolar and nucleoplasmic RNA polymerases，*Proceedings of the National Academy of Science* 65（3）：675-82，March 1970.）

程。RNA 聚合酶 III 催化合成 tRNA、5S rRNA 的前体以及其他小分子 RNA。

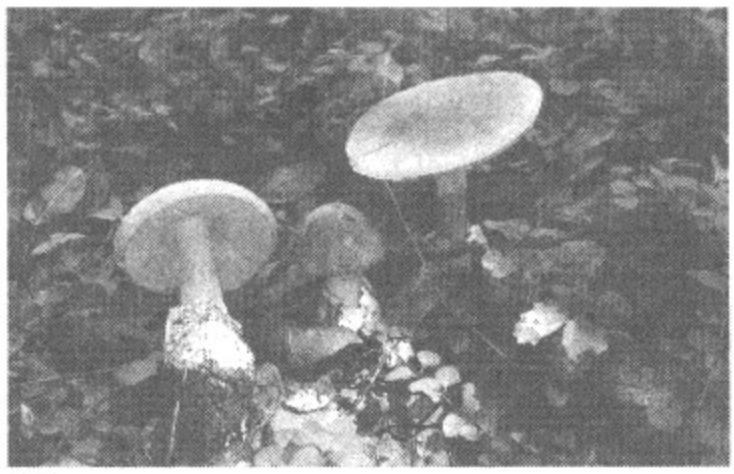
表 10.1　真核生物 RNA 聚合酶的作用

RNA 聚合酶	合成的细胞 RNA	成熟的 RNA（脊椎动物）
I	大的 rRNA 前体	28S、18S 和 5.8S rRNA
II	核内不均一 RNA	mRNA
	核内小 RNA	snRNA
III	5S rRNA 前体	5S rRNA
	tRNA 前体	tRNA
	U6 snRNA（前体?）	U6 snRNA
	7SL RNA（前体?）	7SL RNA
	7SK RNA（前体?）	7SK RNA

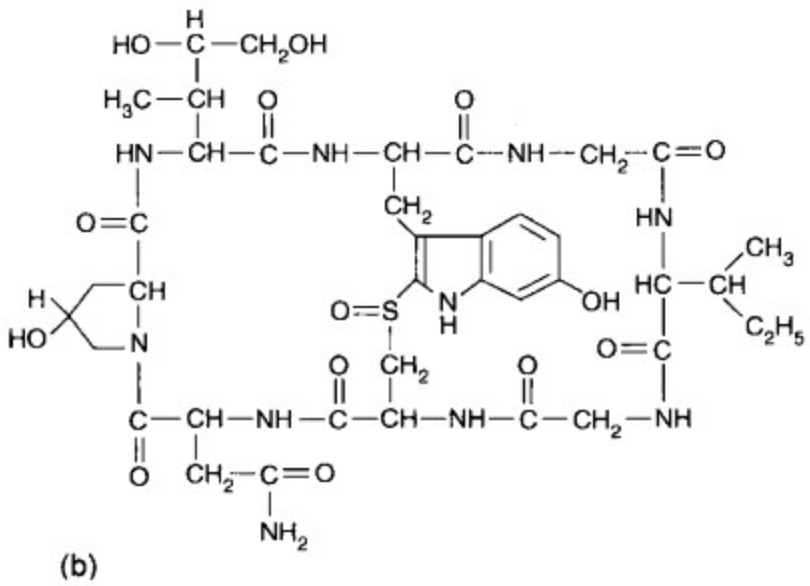
在基因克隆和真核体外转录系统建立之前，我们已经获得有关聚合酶转录分工的证据。在本节，

我们主要分析有关 RNA 聚合酶 III 转录 tRNA 和 5S rRNA 基因的早期证据。

1974 年，Roeder 及同事借助一种称为 α-鹅膏蕈碱（α-amanitin）的毒素完成了这项工作。这种强毒性物质是在伞形毒菌属（*Amanita*）的一些毒蘑中发现的（图 10.3a），包括被称为“死亡之帽”的鬼笔鹅膏菌（*A. pballoides*）和因色泽纯白、致死毒性而被称为“死亡天使”的伞形毒蕈（*A. bisporigera*），这两种毒蘑对许多没有经验的采蘑者是致命的。α-鹅膏蕈碱对三类 RNA 聚合酶有不同的影响，低浓度时，α-鹅膏蕈碱可完全抑制 RNA 聚合酶 II 的活性，而对聚合酶 I 和聚合酶 III 则没有影响；当浓度提高 1000 倍时，大多数真核生物的聚合酶 III 活性都会受到影响（图 10.4）。



(a)



(b)

图 10.3　α-鹅膏蕈碱。（a）鬼笔鹅膏菌（“死亡之帽”），一种能产生 α-鹅膏蕈碱的致死性毒蘑。（b）α-鹅膏蕈碱的结构。（*Source*：（a）Arora, D. *Mushrooms Demystified* 2e, 1986, Plate 50 (Ten Speed Press).）

该实验的步骤是，在 α-鹅膏蕈碱浓度逐渐增加的条件下孵育小鼠细胞核，然后电泳转录产物，观察毒素对小 RNA 分子合成的影响。图 10.5 显示，

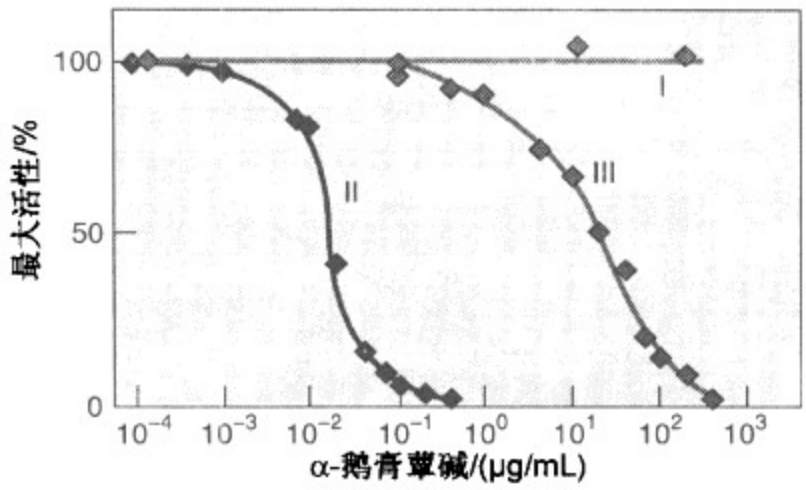


图 10.4　纯化的 RNA 聚合酶对 α-鹅膏蕈碱的敏感性。Weinmann 和 Roeder 分析了 RNA 聚合酶 I（绿色）、RNA 聚合酶 II（蓝色）、RNA 聚合酶 III（红色）在 α-鹅膏蕈碱浓度逐渐升高的条件下，其活性发生变化情况。在 α-鹅膏蕈碱浓度为 0.02μg/mL时，50%的 RNA 聚合酶 II 活性受到抑制，而只有当 α-鹅膏蕈碱浓度达到 20μg/mL 时，才能抑制 RNA 聚合酶 III 的 50%活性；即使 α-鹅膏蕈碱浓度达到 200μg/mL，RNA 聚合酶 I 的活性也不会受影响。（*Source*: Adapted from R. Weinmann and R. G. Roeder, Role of DNA-Dependent RNA polymerase Ⅲ in the transcription of the tRNA and 5S RNA genes, *Proceedings of the National Academy of Science USA* 71 (5): 1790-4, May 1974.）

高浓度的 α-鹅膏蕈碱抑制了 5S rRNA 和 4S tRNA 前体的合成，而且，抑制模式恰好与 RNA 聚合酶 III 被抑制的模式相匹配：在 α-鹅膏蕈碱浓度达到 10 μg/mL 时，酶的 50%活性以及前体 50%的合成量均被抑制了。因此，这些数据支持 RNA 聚合酶 III 是合成这两类 RNA 酶的假设（实际上，RNA 聚合酶 III 也能合成比 5S rRNA 稍大的 5S rRNA 前体，但该实验未能将成熟的 5S rRNA 与其前体区分开来）。RNA 聚合酶 III 也能合成其他多种胞内小分子 RNA 和病毒 RNA，其中包括参与 RNA 剪接（第 14 章）的小 RNA 分子 **U6 snRNA**、在分泌蛋白（secreted protein）合成过程中参与信号肽（signal peptide）识别的小 RNA 分子 **7SL RNA**、与 II 类转录延伸因子 P-TEFb 结合并抑制其功能的核内小 RNA 分子 7SK RNA、腺病毒的 VA（virus-associated）RNA 以及 Epstein-Barr 病毒的 EBER2 RNA。

通过类似的实验，确认了由 RNA 聚合酶 I 和 RNA 聚合酶 II 转录的基因。但对这些结果的解析并不容易，通过更为确定的体外实验进一步证实了这些研究成果。

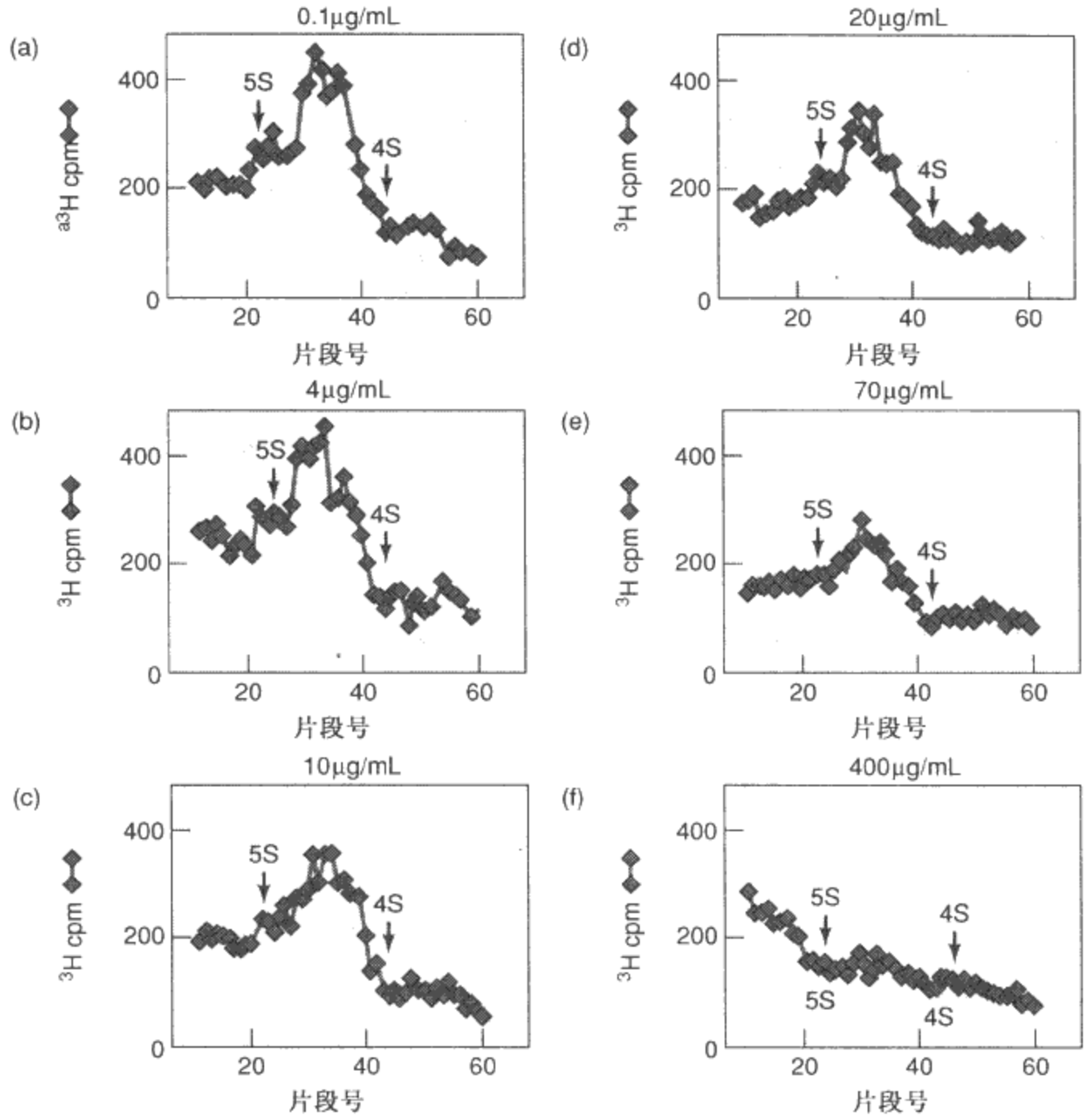


图 10.5　α-鹅膏蕈碱对小 RNA 合成的影响。Weinmann 和 Roeder 利用分离的细胞核进行小分子标记 RNA 的合成，其反应条件是不断增加 α-鹅膏蕈碱的浓度（每个实验所用浓度标注在对应图的上方）。通过离心，从细胞核中分离出标记小分子 RNA，离心后存在于上清液中。对上清进行 PAGE 电泳，测定凝胶中标记 RNA 分子的放射性（红色）。以 5S rRNA 和 4S tRNA 为标准，在同一块凝胶上进行电泳。α-鹅膏蕈碱对 5S rRNA 前体和 4S tRNA 前体合成的影响与图 10.4 中该毒素对 RNA 聚合酶 III 活性的抑制效应是一致的。（*Source*：Adapted from R. Weinmann and R. G. Roeder, Role of DNA-Dependent RNA polymerase III in the transcription of the tRNA and 5S RNA genes, *Proceedings of the National Academy of Science USA* 71 (5)：1790-4, May 1974.）

小结　三类细胞核 RNA 聚合酶在转录过程中行使不同的功能。RNA 聚合酶 I 负责合成较大的 rRNA（在脊椎动物中为 5.8S、18S 和 28S rRNA）前体；RNA 聚合酶 II 催化合成 mRNA 的前体 hnRNA；RNA 聚合酶 III 催化合成 5S rRNA、tRNA 的前体以及其他几种胞内小分子 RNA 和病毒 RNA。

RNA 聚合酶亚基结构

1971 年，Pierre Chambon 小组与 Rutter 小组分别首次独立地报道了真核生物 RNA 聚合酶（聚合酶 II）的第一个亚基结构，但都不全面。顺便提一下，Chambon 将三类聚合酶分别命名为聚合酶 A、B 和 C，而不是聚合酶 I、II 和 III，但是，大多数研究者仍然采用 Roeder 和 Rutter 的命名，即聚合酶 I、II 和 III。1975 年，Roeder 及同事首次对三类细胞核 RNA 聚合酶（来自小鼠培养细胞）比较完整的亚基结构进行了比较。目前，我们已经获得了有关各种真核生物三类 RNA 聚合酶的详细结构信息。

RNA 聚合酶 II 的结构　对真核生物 RNA 聚合酶这样复杂的酶而言，在与活性聚合酶共纯化的多

肽中，我们很难区分哪些多肽是聚合酶的真正亚基，哪些是与酶紧密结合的无关杂蛋白质。解决这个问题的途径之一是，将推定的聚合酶亚基纯化，再重新组装成有活性的聚合酶，从而分析判断哪个亚基是聚合酶活性重构过程中所必需的，进而推断该亚基是否为聚合酶的真正亚基。尽管该策略在研究原核生物聚合酶时是十分完美的，但到目前为止，还没有实现将真核生物细胞核 RNA 聚合酶各个分离的亚基重新组装成有活性的聚合酶，因此，我们必须另辟蹊径。

解决该问题的另一途径是克隆聚合酶所有推定亚基的基因，然后使其突变，从而鉴定哪些亚基是聚合酶活性所必需的组分。利用这种方法已经完成了对酿酒酵母（*Saccharomyces cerevisiae*）聚合酶 II 的亚基组成分析。一些研究人员用传统方法获得高度纯化的聚合酶 II，确定了 10 个亚基，后来，一些研究者又发现了 2 个在早期分析中未被发现的亚基。因此，目前认为酵母聚合酶 II 由 12 个亚基组成。编码这 12 个亚基的基因均已被克隆测序，由此获得了基因产物的氨基酸序列。此外，还对这些基因进行了系统突变实验，分析这些突变对聚合酶 II 活性的影响。

表 10.2 列举了人和酵母聚合酶 II 的 12 个亚基及其分子质量等特性。每个多肽都是由单基因编码，这些聚合酶亚基名称（如 Rpb1 等）是根据相应编码基因的名称（如 *RPB1* 等）而命名的，值得注意的是，Chambon 命名法中的 *RPB* 代表 RNA 聚合酶 B（或 II）。

表 10.2　人与酵母 RNA 聚合酶 II 的亚基

亚基	酵母基因	酵母蛋白/kDa	特征
hRPB1	*RPB1*	192	包含有 CTD；结合 DNA；参与起始位点的选择；与 β′ 同源
hRPB2	*RPB2*	139	包含有活性位点；参与起始位点的选择和延伸速率；与 β 同源
hRPB3	*RPB3*	35	与 Rpb11 共同发挥功能，与原核生物 RNA 聚合酶的 α 二聚体同源
hRPB4	*RPB4*	25	与 Rpb7 组成亚复合体，参与胁迫应答
hRPB5	*RPB5*	25	存在于 pol I、II 和 III 中；转录激活子的靶位点
hRPB6	*RPB6*	18	存在于 pol I、II 和 III 中；装配与稳定功能
hRPB7	*RPB7*	19	与 Rpb4 形成亚复合体，在静止阶段优先结合
hRPB8	*RPB8*	17	存在于 pol I、II 和 III 中；有寡核苷酸/寡糖结合域
hRPB9	*RPB9*	14	含有参与延伸的锌带基序；选择起始位点的功能
hRPB10	*RPB10*	8	存在于 pol I、II 和 III 中
hRPB11	*RPB11*	14	与 Rpb3 共同发挥功能，与原核生物 RNA 聚合酶的 α 二聚体同源
hRPB12	*RPB12*	8	存在于 pol I、II 和 III 中

Source: Reprinted with permission from the *Annual Review of Genetics*, Volume 34, © 2000 by Annual Reviews. www. annualreviews. org.

相比之下，聚合酶 I 和聚合酶 III 的结构又是怎样的呢？首先，所有真核生物的 RNA 聚合酶结构要比细菌聚合酶结构复杂得多；其次，三类真核生物 RNA 聚合酶在结构上都有一定的相似之处，每一类 RNA 聚合酶都含有 2 个大亚基（分子质量超过 100kDa）以及一些小亚基。就这一点而言，它们与原核生物 RNA 核心聚合酶在结构上具有相似之处，原核生物的核心聚合酶由 2 个高分子质量的亚基（β 亚基和 β′亚基）和 3 个低分子质量的亚基（2 个 α′亚基和 1 个 ω 亚基）组成。事实上，正如我们将在本章后面内容了解到的那样，原核生物 RNA 核心聚合酶的 3 个亚基与三类真核生物 RNA 聚合酶的 3 个亚基之间存在明显的进化渊源。换而言之，三类真核生物聚合酶与原核生物 RNA 核心聚合酶之间存在着关联，而且 3 类真核生物 RAN 聚合酶彼此之间也存在关联。

表 10.2 提供的第三个信息是有些亚基是酵母三类细胞核 RNA 聚合酶所共有的，事实上，存在着 5 个共有亚基（common subunit）。在 RNA 聚合酶 II 结构中，这些共有亚基分别被称为 Rpb5、Rpb6、Rpb8、Rpb10 和 Rpb12，已在表 10.2 的右栏中标注。

Richard Young 及同事已经确证，最初认定的 10 个多肽确实是 RNA 聚合酶 II 的亚基，或至少是

紧密结合的杂蛋白质。研究者用表位附加法（epitope tagging）（见图 10.6），通过基因工程手段，在酵母 RNA 聚合酶 II 的一个亚基（Rpb3）上附加一小的外源抗原决定簇，然后将该融合基因导入 *Rpb3* 基因功能缺失的酵母细胞，用^{35}S 或^{32}P 标记细胞蛋白质，通过外源表位的特异抗体沉淀全酶。免疫共沉淀后，用 DS-PAGE 从沉淀蛋白中分离出标记的多肽，放射自显影检测。图 10.7a 为实验结果。通过这个一步纯化法获得了纯化的明显含有 10 个亚基的 RNA 聚合酶 II，还能观察到一些小肽，而且也存在于没有附加表位的野生型 RNA 聚合酶 II 对照组中。说明这些小肽与 RNA 聚合酶不相干。图 10.7b 是 Roger Kornberg 及同事对相同聚合酶进行 SDS-PAGE 分析的实验结果，共识别出 12 个亚基，其中 Rpb11 与 Rpb9 以及 Rpb12 与 Rpb10 分别存在着共电泳现象，所以，早期的研究未能识别出 Rpb11 和 Rpb12 这 2 个亚基。

由于 Young 及同事已经知道了所有最初 10 个亚基的氨基酸组成，因此用^{35}S-甲硫氨酸标记每条多肽后就可以很好的估算每个亚基的化学计量，表 10.3 列举了各亚基的化学计量。图 10.7a 也表明有 2 个 RNA 聚合酶 II 亚基 Rpb1 和 Rpb6 发生了磷酸化修饰，因为这 2 个亚基可被［γ-^{32}P］ATP 标记。Rpb2 也可被磷酸化修饰，但磷酸化程度较低，以至于在图 10.7a 中观察不到这种修饰作用。

表 10.3　酵母 RNA 聚合酶 II 的亚基

亚基	SDS-PAGE 迁移率/kDa	蛋白质分子质量/kDa	化学计量	缺失表型
RPB1	220	190	1.1	致死
RPB2	150	140	1.0	致死
RPB3	45	35	2.1	致死
RPB4	32	25	0.5	条件致死
RPB5	27	25	2.0	致死
RPB6	23	18	0.9	致死
RPB7	17	19	0.5	致死
RPB8	14	17	0.8	致死
RPB9	13	14	2.0	条件致死
RPB10	10	8.3	0.9	致死
RPB11	13	14	1.0	致死
RPB12	10	7.7	1.0	致死

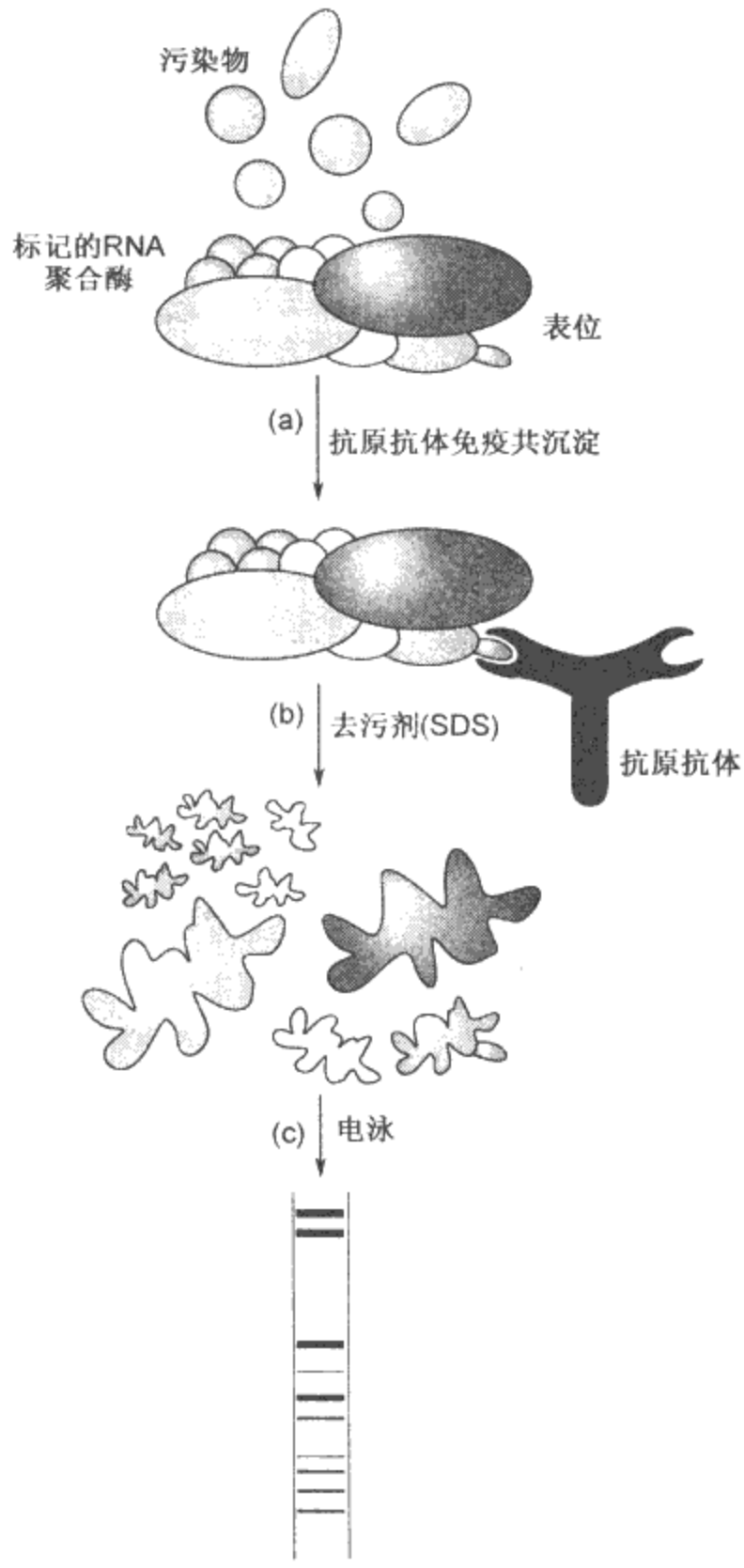


图 10.6　表位附加法的原理。通过基因工程方法，将一外源功能域（表位，红色）重组到酵母 RNA 聚合酶 II 的 Rpb3 亚基上，其他正常的亚基仍能够与改变了的 Rpb3 组装成有功能活性的聚合酶。在含有放射性标记氨基酸的培养基中培养酵母细胞，使该聚合酶被标记。（a）添加识别表位的抗体，通过免疫共沉淀反应，使 RNA 聚合酶与其他杂蛋白（灰色）分离，只经过一步纯化即可获得高度纯度的 RNA 聚合酶。（b）加入强去污剂 SDS，使纯化聚合酶中的各个亚基分离变性。（c）凝胶电泳变性检测变性的聚合酶亚基，获得图下方的电泳图谱。

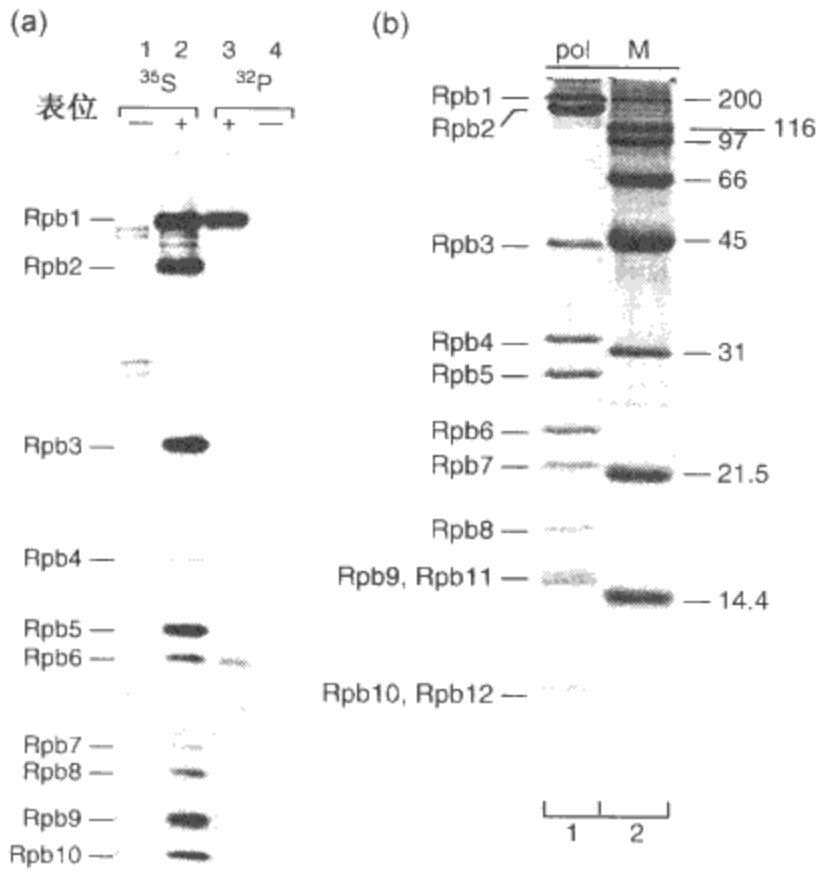


图 10.7　酵母 RNA 聚合酶 II 的亚基结构。（a）用表位附加法获得了约包含有 10 个亚基的结构。Young 及同事通过融合基因（编码抗原决定簇标签和 RPB3 亚基）取代正常 *RPB3* 基因的方法，使 RPB3 亚基携带有额外的表位。然后用［^{35}S］甲硫氨酸标记聚合酶的所有亚基，或用［γ-^{32}P］ATP 只标记磷酸化亚基。利用表位特异抗体进行免疫共沉淀分离标记的蛋白质，凝胶电泳检测。泳道 1：从无表位的野生型酵母菌株细胞中分离的^{35}S 标记的蛋白质；泳道 2：从带有表位的酵母菌株细胞中分离的^{35}S 标记的蛋白质；泳道 3：从带有表位的酵母菌株细胞中分离的^{32}P 标记的蛋白质；泳道 4：从野生型酵母菌株细胞中分离的^{32}P 标记的蛋白质。图版左侧标注了聚合酶 II 的各个亚基。（b）经包括免疫共沉淀的多步纯化，获得了约由 12 个亚基组成的聚合酶结构。Kornberg 及同事利用免疫共沉淀方法分离纯化了酵母的 RNA 聚合酶 II，然后进行 SDS-PAGE 凝胶电泳（泳道 1），泳道 2 为分子质量标记。图右侧标注了分子质量大小，左侧标注了亚基组分。Rpb9 与 Rpb11 及 Rpb10 与 Rpb12 的迁移速率几乎相同。（*Source*:(a) Kolodziej, P. A., N. Woychik, S. -M. Liao, and R. Young, RNA polymerase II subunit composition, stoichiometry, and phosphorylation. *Molcular and Cellular Biology*. 10 (May1990) p. 1917, f. 2. American Society for Microbiology. (b) Sayre, M. H., H. Tschochner, and R. D., Kornberg, Reconstitution of transcription with five purified initiation factors and RNA polymerase II from *Saccharomyces cerevisiae*. *J. Bio. Chem*. 267 (15 Nov 1992) p. 23379, f. 3b. American Society for Biochemistry and Molecular Biology.)

Young 将 RNA 聚合酶 II 最初的 10 个亚基分为三组：核心亚基（core subunit），在结构和功能上与细菌 RNA 聚合酶相对应；共有亚基（common subunit），至少在酵母中是三类细胞核 RNA 聚合酶共有的；非必需亚基（nonessential subunit），是 RNA 聚合酶活性条件非必需亚基。

核心亚基　包括 Rpb1、Rpb2 和 Rpb3，都是 RNA 聚合酶活性所绝对必需的，分别与 *E. coli* RNA 聚合酶的 β'、β 和 α 亚基同源。

那么，它们之间的功能关系如何？我们已经了解到（第 6 章），*E. coli* RNA 聚合酶的 β' 亚基结合 DNA，Rpb1 的作用也是结合 DNA。我们还了解到，*E. coli* RNA 聚合酶的 β 亚基位于或靠近酶的核苷酸连接活性位点处。Andre Sentenac 及同事用相同的实验方案确认 Rpb2 也位于或邻近 RNA 聚合酶 II 的活性位点处。在所有三类细胞核 RNA 聚合酶中，其第二大亚基间具有功能相似性，与原核生物 RNA 聚合酶也存在着功能相似性，这种功能相似性可由结构上的相似性得以反映，基因序列揭示了亚基间的结构相似性。

尽管 Rpb3 与 *E. coli* 的 α 亚基并不是十分相似，但存在一个长 20 个氨基酸的高度相似区域。此外，这两个亚基大小相近，化学计量相同，每个全酶含有 2 个该亚基的单体。而且，*RPB3* 突变及 *E. coli* α 亚基突变会引起相同的聚合酶组装缺陷表型。所有的这些因素都表明，Rpb3 与 *E. coli* 的 α 亚基同源。

共有亚基　包括 Rpb5、Rpb6、Rpb8、Rpb10 和 Rpb12 共 5 个亚基，是酵母细胞三类 RNA 聚合酶共有的亚基组分，目前对这些亚基的功能还知之甚少，但是，既然这些亚基存在于所有三类 RNA 聚合酶中，表明它们在转录过程具有极其重要的功能。

非必需亚基　1989 年，Woychik 和 Young 分离出了 *rpb4* 酵母突变菌株，该突变体缺失编码 Rpb4 亚基的基因，正如所料，该酵母菌株的 RNA 聚合酶 II 没有 Rpb4 蛋白。而且，研究者发现，尽管 Rpb7 亚基的编码基因未被改变而是保持野生型状态，但是该酵母菌株的 RNA 聚合酶 II 同样缺失 Rpb7 蛋白。这一发现暗示，Rpb4 具有将 Rpb7 锚定到聚合酶上的功能，两者较容易从 RNA 聚合酶 II 上脱离下来，而且解离下来的 Rpb4 和 Rpb7 又通常低于化学计量。更有一些证据表明，这两个亚基能够在不同的 RNA 聚合酶 II 间来回穿梭。

Roger Kornberg 及同事用 *rpb4* 酵母突变菌株的 RNA 聚合酶 II（pol IIΔ4/7）进行体外转录实验，发现该突变体的 RNA 聚合酶能够很好地转录缺失启动子的 DNA 分子，但不能在真正的启动子处起始转录。换而言之，这种突变的 RNA 聚合酶具有

正常的转录延伸和转录终止功能，但失去了转录起始功能。需要强调的是，Rpb7 是 RNA 聚合酶 II 的必需亚基，不可能在 pol II A4/7 突变体中完全缺失。

Rpb7 和 Rpb11 虽然没有被归纳到上述三组亚基中，但它们是聚合酶活性所绝对必需的。*Rpb7* 和 *Rpb11* 突变是致死的。

小结 RNA 聚合酶 II 中 12 个亚基的编码基因均已被克隆、测序，并进行了突变分析。有 3 个亚基在结构和功能上与细菌 RNA 聚合酶的核心亚基相似；有 5 个亚基是三类细胞核 RNA 聚合酶所共有的；有 2 个亚基并不是聚合酶发挥活性所必需的，至少在 37℃条件下如此；有 2 个亚基不能被归结到上述三组亚基中。将标记核苷酸和磷酸基团掺入 RNA 聚合酶 II 中，能够帮助我们估算这 12 个亚基的化学计量。在这 12 个亚基中，有 2 个亚基是高度磷酸化的，有 1 个亚基是轻度磷酸化的。

Rpb1 亚基的异质性 对 RNA 聚合酶 II 结构的早期研究表明，其最大亚基具有一定的异质性（heterogeneity），图 10.8 示例了小鼠浆细胞瘤（plasmacytoma）中 RNA 聚合酶 II 的异质性现象。电泳后观察发现，凝胶上部的 3 个多肽 IIo、IIa 和 IIb 在数量上略少于多肽 IIc。这 3 个多肽似乎彼此关联，实际上，其中的 2 个多肽可能是第三个多肽的衍生物，但哪个是母体？哪两个是衍生物？对酵母 *RPB1* 基因测序后，推导其多肽产物的分子质量为 210kDa。所以，分子质量接近 210kDa 的 IIa 可能是母体。

而且，氨基酸测序结果表明，IIb 亚基缺失一段由 7 个氨基酸残基组成的串联重复序列，该七肽聚体的保守一致性序列为 Tyr-Ser-Pro-Thr-Ser-Pro-Ser。因为这段序列存在于 IIa 亚基的羧基末端，所以称之为碳末端结构域（carboxyl-terminal domain，CTD）。CTD 的抗体易与 IIa 亚基发生反应，但不会与 IIb 发生相互作用，这就进一步证实了 IIb 缺失 CTD 的结论。对这种异质性的可能解释是，蛋白质水解酶切除了 CTD，将 IIa 转化为 IIb。由于在体内没有观察到 IIb 的存在，所以，IIb 可能是在纯化 RNA 聚合酶 II 过程中人为造成蛋白质分解而形成的产物。事实上，CTD 的序列特征表明，CTD 不但不会折叠成紧密结构，反而可能是伸展的，易受蛋白质水解酶的作用而降解。

那么 IIo 亚基的情况又如何呢？该亚基比 IIa 大，显然不是由 IIa 通过水解形成的，而可能是 IIa 的磷酸化产物，该结论获得了若干证据的支持。首先，蛋白质的磷酸化修饰主要发生在丝氨酸、苏氨酸的羟基，有时是酪氨酸的羟基上，而 CTD 主要由这 3 个氨基酸组成的事实表明，该结构域很可能会发生磷酸化修饰。事实上，如果将 IIo 亚基与磷酸酶共同孵育，IIo 亚基上的磷酸基团可被去除而转化为 IIa。此外，在 IIo 亚基的 CTD 中，第 2 位和第 5 位丝氨酸残基被磷酸化。

那么，IIo 与 IIa 之间分子质量的差异仅仅就是由于磷酸基团的存在与否？事实显然并非如此，即使在 CTD 重复 52 次的哺乳动物 RNA 聚合酶 II 中也没有发现大量磷酸基团的存在，所以，我们必须另寻能够解释 IIo 电泳迁移率低的原因。或许是 CTD 磷酸化后引起 IIo 构象发生改变，从而使 IIo 的电泳迁移率变缓，表现出比其实际大小大得多的行为。但是，即使蛋白质发生了变性，也会维持已有的构象。图 10.9 描述了 IIo、IIa 和 IIb 这 3 个亚基之间可能的关系。

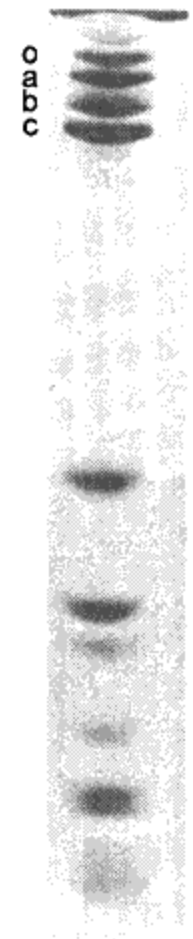


图 10.8 小鼠浆细胞瘤 RNA 聚合酶 II 的部分亚基结构。图左侧的字母 o、a、b 代表最大亚基的 3 种不同形式，此最大亚基与酵母 RNA 聚合酶 II 的 Rpb1 亚基相对应，亚基 c 与酵母 RNA 聚合酶 II 的 Rpb2 亚基相对应。（*Source*：Sklar，V. E. F.，L. B. Schwartz，and R. G. Roeder，Distinct molecular structures of nuclear class I，II，and III DNA-dependent RNA polymerases. *Proceedings of the National Academy of Sciences USA* 72（Jan 1975）p. 350，f. 2C.）

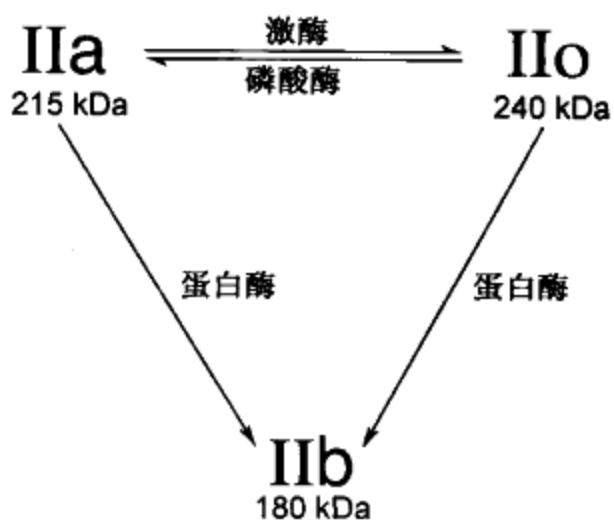


图 10.9 RNA 聚合酶 II 的最大亚基的不同形式之间可能存在的关系。

细胞内 Rpb1 亚基（IIo 和 IIa）存在两种形式的事实暗示 RNA 聚合酶 II 以两种不同的形式存在于细胞内，每种形式的 RNA 聚合酶 II 包含有一种形式的 Rpb1 亚基，分别称之为 **RNA 聚合酶 IIO 和 RNA 聚合酶 IIA**，含有 IIb 亚基的非磷酸化 RNA 聚合酶 II 则被称为 **RNA 聚合酶 IIB**。

在细胞内，聚合酶 IIO 和 IIA 的作用相同还是截然不同？有充分的证据表明，聚合酶 IIA（RNA 聚合酶 II 的非磷酸化形式）的作用是起始与启动子的结合，而聚合酶 IIO（CTD 被磷酸化修饰）则主要执行延伸功能。由此可见，CTD 的磷酸化修饰伴随着转录由起始向延伸的过渡。我们将在第 11 章讨论该假说的相关证据。

小结　酵母 *RPB1* 基因的原初产物为 IIa 亚基。IIa 亚基碳末端功能域（CTD）由 7 个氨基酸组成的重复序列构成，在体外，CTD 可被蛋白酶切除而使 IIa 转化为 IIb。此外，CTD 上的 2 个丝氨酸可被磷酸化修饰而使 IIa 转化为 IIo。含有 IIa 亚基的 RNA 聚合酶 IIA 负责与启动子的结合，含有 IIo 亚基的 RNA 聚合酶 IIO 负责转录的延伸。

RNA 聚合酶 II 的三维结构　正如我们在第 9 章所了解到的那样，获得蛋白质结构信息的传统方法是进行 X 射线晶体学分析，利用该方法对嗜热水生菌（*Thermus aquaticus*）和 T7 噬菌体的 RNA 聚合酶的结构进行了解析。但直到 1999 年，还很难获得高质量的 RNA 聚合酶 II 晶体用于 X 射线晶体学研究。主要的问题是，在结晶过程中有些酶因丢失 Rpb4 和 Rpb7 亚基而导致酶的异质性（异质性蛋白质的混合物不易形成结晶体）。Roger Kornberg 及同事利用缺失 Rbp4（同时也缺失 Rpb7）亚基的酵母 RNA 聚合酶 II 突变体（pol II△4/7）成功克服了蛋白质的异质性问题，尽管不能在启动子处起始转录，但该酶具有转录延伸的功能。因此，pol II△4/7 完全可以作为分析延伸复合体结构的模型，2001 年，以 pol II△4/7 晶体为研究对象，获得了分辨率为 2.8Å 的 X 射线晶体图像。

图 10.10 显示的是酵母 RNA 聚合酶 II 模型的立体图像，每个亚基用不同的色彩标注区分，图右上方的简图注明了各个亚基间的相对位置。此酶最突出的特征是催化活性位点位于一个很深的 DNA 结合裂隙中，裂隙底部结合有一个 Mg^{2+}。张开的 DNA 结合裂隙如同钳状结构，上半部分由 Rpb1 和 Rpb9 亚基组成，下半部分由 Rpb5 亚基形成。

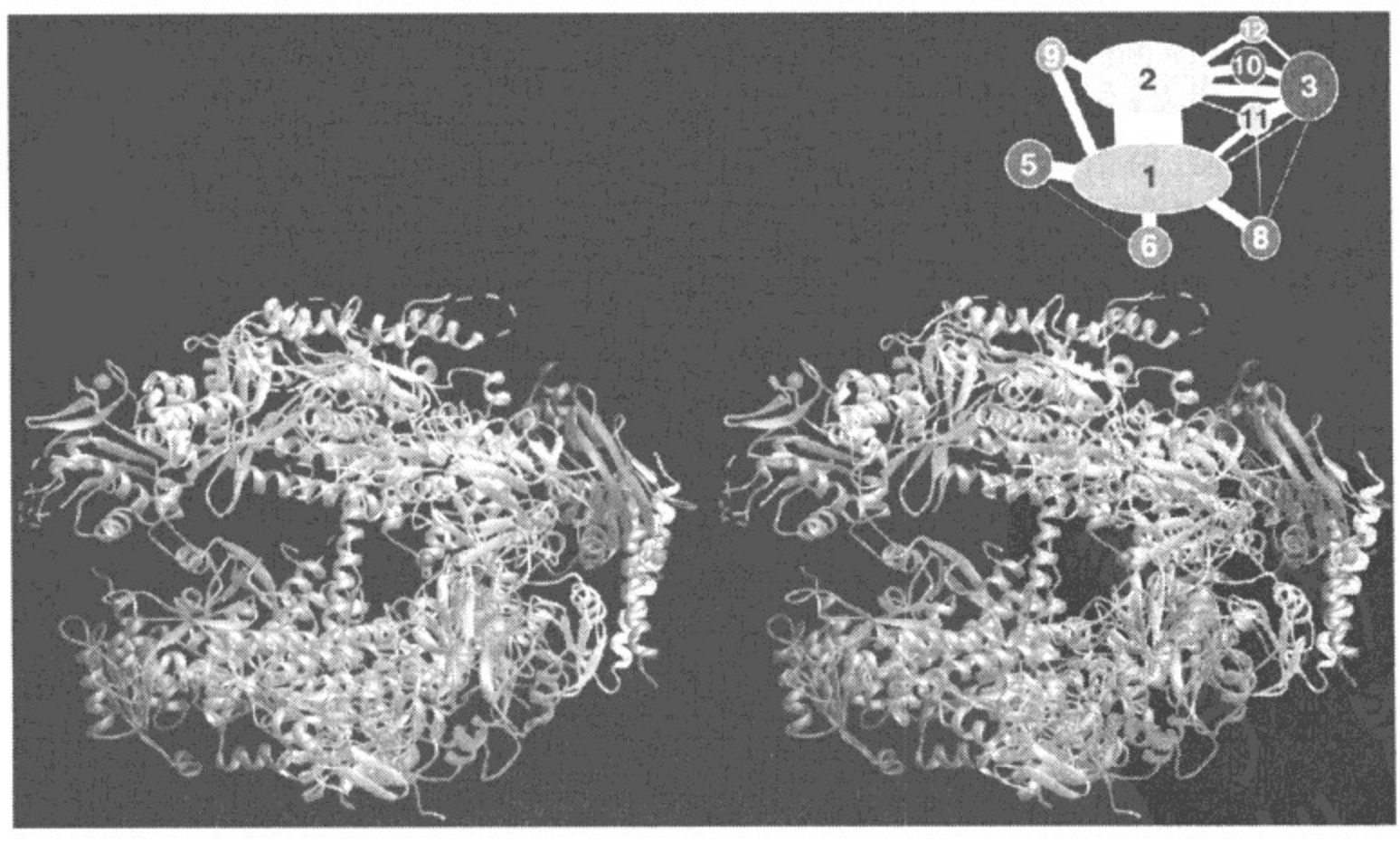


图 10.10　酵母 RNA 聚合酶 II 的晶体结构。图下方的立体图像显示了酵母 RNA 聚合酶 II 中的 10 个亚基（缺少 Rpb4 和 Rpb7 亚基），亚基的颜色与图右上方结构示意图中的相同。示意图中白色线条的粗细代表两个亚基之间作用的强度，在立体图像中，紫红色球体代表位于活性中心的金属离子，Zn^{2+} 用蓝色球体表示。（*Source*：Cramer，et al.，*Science* 292：p. 1864.）

Kornberg 及同事研究低分辨率的晶体结构图像时发现，DNA 模板位于酶的裂隙之中，新结构图像证实了这一假说。研究显示，酶的裂隙由碱性氨基酸残基构成，而酶的表面几乎全部由酸性氨基酸构成（图 10.11，前视及俯视），位于裂隙中的碱性氨基酸残基有助于酶与酸性 DNA 模板的结合。

对所有单亚基 RNA 聚合酶及 DNA 聚合酶的结构研究发现，活性中心结合有 2 个金属离子，由此提出了依赖于 2 个金属离子的催化机制。先前对酵母 RNA 聚合酶 II 晶体结构进行研究时，在催化中心只发现了 1 个 Mg^{2+}，然而，高分辨率的晶体结构图像显示，在催化中心确实存在 2 个 Mg^{2+}，尽管其中的一个信号较弱。Romberg 及同事推测信号强的 Mg^{2+}（*metal A*）与活性中心结合紧密，信号弱的 Mg^{2+}（*metal B*）与活性中心结合松散，可能参与结合核苷酸底物。金属离子 A 与 Rpb1 亚基的 3 个保守天冬氨酸残基（D481、D483 和 D485）紧密结合，金属离子 B 被 3 个酸性氨基酸残基所包围，即 Rpb1 亚基的 D481、Rpb2 亚基的 E836 和 E837。在晶体结构中，这些氨基酸残基与金属离子 B 相距较远，但在催化反应过程中能够靠近并结合金属离子 B，在催化中心形成适当的构象以促进聚合反应。

如同我们在第 6 章学过的，聚合酶能够发生暂停并向后移动，将初生 RNA 链的 3′端排出酶外。酶底部的小孔称为**孔 1**（pore 1），是 RNA 产物从酶中排出的位点（图 10.11，俯视图），这一假说得到了新晶体结构图像的支持。孔 1 通向酶底部的一漏斗状结构，并被同源的细菌 RNA 聚合酶所证实，位于漏斗内的 β′亚基与排出的 RNA 链末端发生交联。此外，能够引发外突 DNA 发生切割的延伸因子 TFIIS（第 11 章）显然结合在漏斗状内，这一发现与下列事实相吻合，即能够与 TFIIS 发生相互作用的 Rpb1 亚基上的氨基酸残基位于漏斗的外缘。

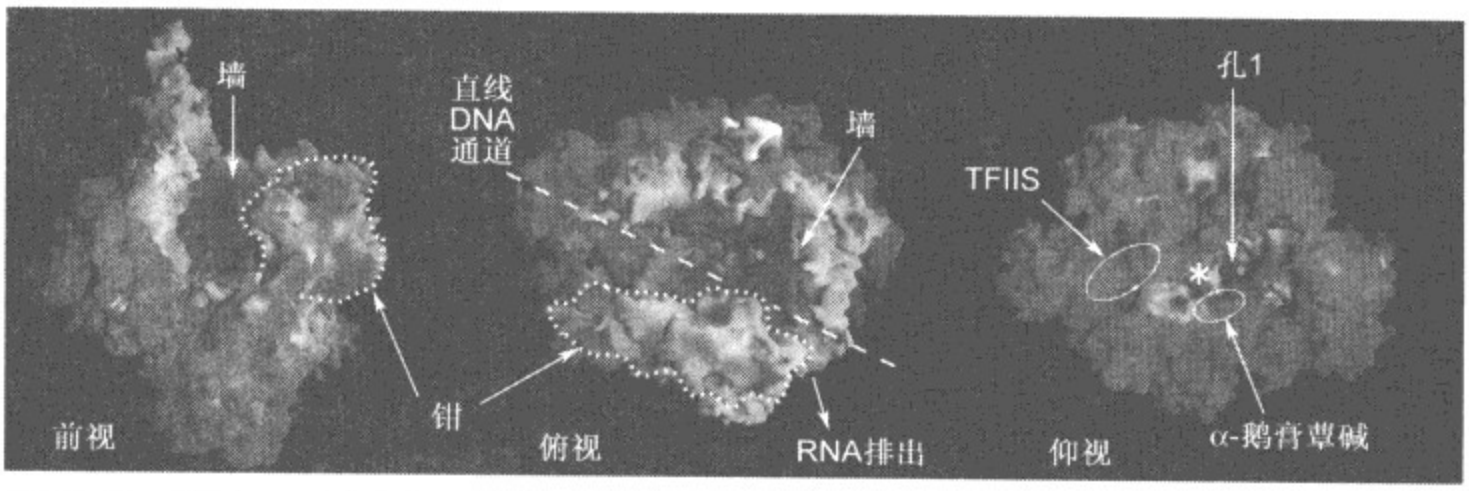


图 10.11　酵母 RNA 聚合酶 II 的表面结构特征。在 3 种视图中，其表面静电势分别用红色表示负（酸性）、蓝色表示正（碱性）、白色表示中性。在俯视或仰视图中，可以观察到用粉红色小球体表示的、位于活性中心的金属离子，钳、阻隔墙等各种特征均在图中标注。在仰视图中，大椭圆标注的位置是 TFIIS 的结合位点，小椭圆标注的位置是 α-鹅膏蕈碱的结合位点，星号是指与 *E. coli* RNA 聚合酶相似的保守序列的起点，该序列能够与排出的 RNA 链的 3′端发生交联。（*Source*：Cramer，et al.，*Science* 292：p. 1872.）

Kornberg 及同事将 RNA 聚合酶 II 晶体浸泡于不同成分的溶液中，使晶体不断收缩而获得了高分辨率的新晶体结构图像。通过该结晶程序获得了体积渐小的两类晶体，其相应的结构称之为构型 1 和构型 2。除了能够提供高分辨率的结构信息外，Kornberg 及同事还根据这两种结构的异同，观察酶结构中哪些部分能够保持固定不变，哪些部分会变换构型。

钳组件（clamp module）可发生最显著的移动，移动距离可达 14Å。在构型 1 中，钳结构处于相对闭合状态，若不发生弯曲，则很难想象 DNA 可以结合在裂隙中。然而，在构型 2 中，钳结构向后摆动，处于打开状态，如图 10.12 所示能够接纳线性 DNA 分子。该模型推测，在转录起始复合体中，DNA 模板与钳结构的 Rpb9 亚基在酶的一侧相接触，位于酶碱性裂隙中的 DNA 分子靠近酶的活性中心，并在酶的另一侧与鞍状结构相接触，该假说也获得了遗传学证据的支持。DNA 进入酶分子内的可能途径包括 Rpb9 钳、Rpb1 和 Rpb2 裂隙中的氨基酸残基发生突变都会产生致死效应或影响转录的起始。该模型的大部分信息都是正确的，但本章后面引用的最新研究数据显示，酶与 DNA 的结合绝不是酶的钳状结构打开以接纳双链 DNA 分子那样简单。事实上，在与启动子 DNA 发生相互作用时，酶的钳状结构处于闭合状态，这一发现提示，启动子 DNA 在酶的表面发生熔解，只有模板链能够进入裂隙中的活性位点，而非模板链则留在外面。

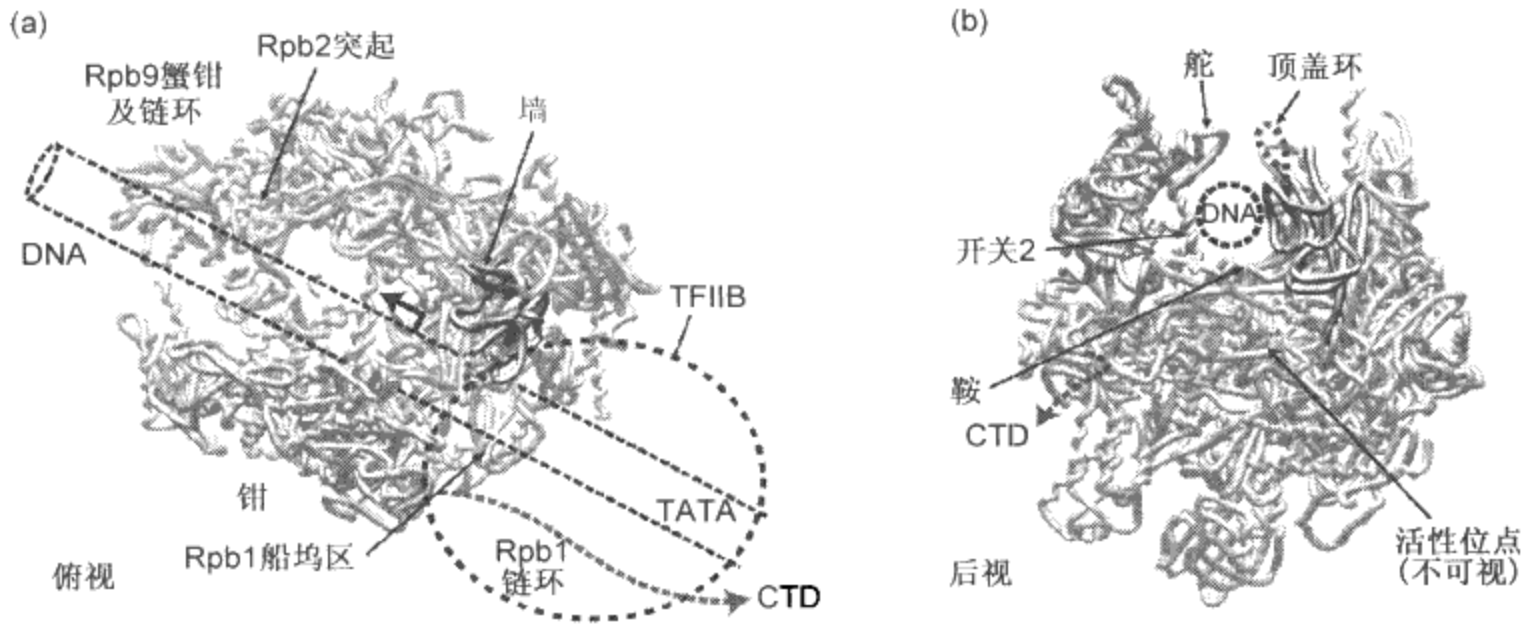


图 10.12　在起始复合体中，含启动子的直线双链 DNA 的结合途径模型。（a）俯视图。钳状结构（红色）处于打开状态，允许直线 DNA 分子进入裂隙。转录起始位点（折线箭头）和 TATA 盒已在图中标注。活性中心的金属离子用紫红色球体表示。Rpb9 钳、接头环（橙色）以及阻隔墙（紫色）也在图中标明。位于 TATA 盒与聚合酶表面之间的 TFIIB 的位置用黑色虚线标出，TFIIB 蛋白与 Rpb1 CTD 的接头重叠。用黄色突出的位点是指此位点发生突变会影响转录的起始。（b）背视图。可观察到位于钳（红色）和阻隔墙（紫色）之间的 DNA 分子。（*Source*：Cramer，et al.，*Science* 292：p. 1873.）

在第 11 章将学习通用转录因子 TFIID 与启动子特异位点 TATA 盒的结合，TATA 盒位于转录起始点上游约 25bp 处。TFIIB 是连接 TFIID 与聚合酶 II 的桥梁因子，将聚合酶Ⅱ定位在正确的起始位置处。图 10.12 所示的模型与下列假说相一致，假定 TATA 盒与转录起始位点间的 DNA 是线性结构，那么，结合在 TATA 盒上的 TFIID 与结合在转录起始位点上的 RNA 聚合酶 II 间的距离与 TFIIB 的大小相吻合。

由于 RNA 聚合酶之间存在着高度同源性，酵母 RNA 聚合酶 II 的结构对于研究其他 RNA 聚合酶是十分有用的。特别是，酵母 RNA 聚合酶 II 与人类 RNA 聚合酶 II 有 53% 的序列一致性。而且，Kornberg 及同事最初在酵母 RNA 聚合酶 II 中鉴定出的 10 个亚基与酵母另两个聚合酶的亚基完全一致或高度同源。另外，酵母 RNA 聚合酶 II 的催化核心在序列和形状上十分类似于细菌聚合酶的催化核心。因此，从酵母 RNA 聚合酶Ⅱ获得的结构信息可以指导分子生物学家深入研究酶的功能，并可开展有关其他 RNA 聚合酶的结构与功能研究。

小结　原子级分辨率的酵母 RNA 聚合酶 II（pol II△4/7）的结构显示，酶内具有一个很深的裂隙，能够接纳两个末端之间的线性 DNA 模板。结合有一个 Mg^{2+} 催化活性中心位于裂隙底部，第二个 Mg^{2+} 以较低的浓度存在，可能参与酶与核苷酸底物的结合过程。线性 DNA 模板与聚合酶复合体的几何结构显示，该复合体具有足够的空间，以利于 TFIID 与启动子 TATA 盒的结合，以及 TFIIB 将聚合酶连接到 TFIID 上，从而将聚合酶 II 定位在正确的起始位点处。

RNA 聚合酶 II 在延伸复合体中的三维结构　在上节内容中介绍了酵母 RNA 聚合酶 II 在原子级分辨率层面上的结构特征，并假定 DNA 模板结合在酶的活性中心。但是，Kornberg 及同事通过研究确认，在延伸复合体中，酵母 RNA 聚合酶 II 既与 DNA 模板结合，又与 RNA 产物结合。该复合体晶体结构图像的分辨率（3.3Å）并不高于聚合酶自身晶体结构图像的分辨率，但却提供了有关酶与 DNA 模板、RNA 产物之间相互作用的大量信息。

为诱导 RNA 聚合酶 II 在没有任何转录因子协助下起始转录，Kornberg 及同事用 3′端带有单链 oligo［dC］尾的 DNA 分子作模板，此模板可使 RNA 聚合酶 II 在与双链区相距 2 或 3 个核苷酸的单链位点处起始 RNA 的转录。而且，实验设计的模板可使聚合酶在缺乏 UTP 的条件下转录合成长度为 14 个核苷酸的 RNA 产物，直到在第一次需要 UTP 时 RNA 聚合酶才会停止转录。这样，在反应体系中就会形成大量同质均一的延伸复合体，同时反应体系中也混有未与 DNA 结合的无活性聚合酶，这些无活性 RNA 聚合酶可通过肝素层析柱去除。作为一种多阴离子表面活性剂，如果酶的裂隙未被

DNA 所占据，肝素就能够与酶的碱性裂隙结合。所以，无活性的聚合酶能够与层析柱上的肝素结合，而具有活性的延伸复合体则能够通过层析柱，这些均质的延伸复合体可被结晶析出。

图 10.13a 显示了延伸复合体及酶自身的晶体结构。除了在延伸复合体中存在核苷酸之外，这两个晶体结构最显著的不同在于钳所处的位置。在酶自身的晶体结构中，钳状结构面向活性位点，处于打开状态。但在延伸复合体的晶体结构中，钳状结构环抱着 DNA 模板和 RNA 产物，从而使酶在转录一个完整基因时具有很高的**进行性**（processive），避免在转录过程中，酶与模板脱离而造成转录的提早终止。

图 10.13b 显示了延伸复合体的近观结构，其中将酶的部分结构切除以利于展现结合在酶裂隙中的核苷酸。该结构具有以下几个明显的特征：首先，9bp 长的 DNA-RNA 杂交分子（由 DNA 模板链与 RNA 产物杂交形成）的对称轴相对于还在被转录的下游 DNA 双螺旋而言具有一定的角度。该弯曲在钳状结构闭合过程中会受到压力，但位于 DNA-RNA 杂交体与下游 DNA 双螺旋之间的单链 DNA 有利于该弯曲的形成。其次，具有催化作用的 Mg^{2+} 位于活性中心，该位点能够将单个核苷酸添加到正在生长的 RNA 链上。该离子就是聚合酶自身晶体结构中的金属 A。最后，**桥梁螺旋**（bridge helix）穿越酶活性位点附近的裂隙。在本节后面内容中，我们将详细探讨桥梁螺旋的结构特征及功能。

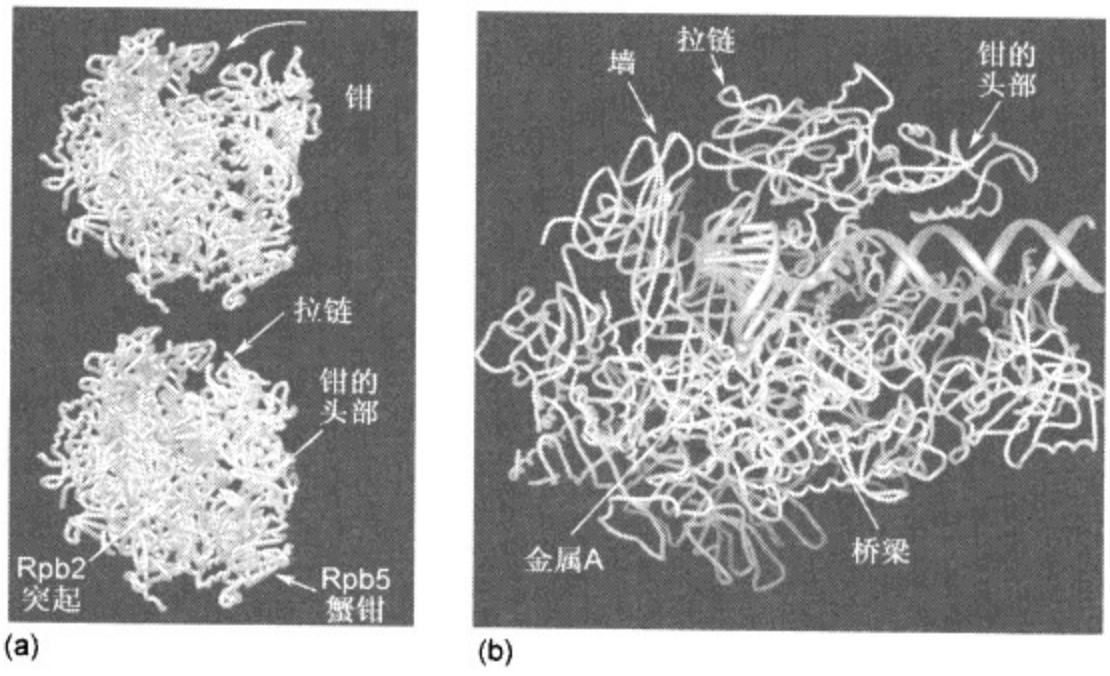


图 10.13　延伸复合体的晶体结构。（a）游离 RNA 聚合酶 II（上）与延伸复合体（下）的晶体结构比较。用黄色突显钳子结构，DNA 模板链、DNA 非模板链、RNA 产物分别用蓝色、绿色和红色表示。（b）延伸复合体的详细图示。颜色标注与（a）相同，活性中心的金属离子用紫红色球体表示，桥梁螺旋用绿色表示。（*Source*：Gnatt，et al.，*Science* 292：p. 1877.）

在延伸复合体中，Mg^{2+}（金属 A）处于能够与连接＋1 和－1 核苷酸（最新添加到正在生长的 RNA 链上的 2 个核苷酸）之间的磷酸基团相结合的位置上（图 10.14a）。延伸复合体缺失了金属 B，可能是由于在最新添加到 RNA 链上的核苷酸释放焦磷酸时，金属 B 伴随着焦磷酸一同与复合体解离。＋1 位核苷酸恰好位于孔 1 的入口处（图 10.14b），表明核苷酸通过孔 1 进入酶的活性位点。事实也是如此，在酶分子中已没有空间允许核苷酸通过其他途径进入酶的活性位点，除非核苷酸和蛋白质的位置发生显著重排，而且，当 RNA 聚合酶发生倒退时，孔 1 恰好位于 RNA 链 3′端从酶中排除的位置处。只有当核苷酸发生错配时 RNA 聚合酶才会出现倒退行为，使错配核苷酸外露，便于结合在孔 1 另一末端通道内的 TFIIS 将此错配核苷酸切除。

图 10.14b 也阐明了盖、舵和拉链这 3 个外突于钳的环结构的可能作用。这 3 个环通过其所在的位置而对转录过程中的几个重要事件产生特定的影响，如转录泡的形成和维持、RNA-DNA 杂交分子的解离。如果 RNA-DNA 杂交分子长度继续延伸超过 9bp 时，舵形蛋白就会产生阻碍作用，促使杂交分子发生解体。

Kornberg 及同事注意到，在延伸复合体中桥梁螺旋为直线结构，但在细菌聚合酶晶体结构中则呈弯曲状态，该弯曲发生在 Thr831 和 Ala832 这两个保守氨基酸残基附近，弯曲的桥梁螺旋可阻止核苷酸与活性位点的结合。由此可推测桥梁螺旋可能在转位时起作用，如图 10.15 所示。他们认为，桥梁螺旋在转位步骤中能够不断改变自己的构象，在直线与弯曲两种构象间来回变换。当桥梁螺旋处于直线构象时，新添加的核苷酸通过位于活性位点下方的孔 1 进入处于打开状态的活性位点，聚合酶便将此核苷酸添加到正在生长的 RNA 链的 3′端，填充了直线型桥梁螺旋与 RNA 链 3′端之间的间隙。伴随着转位，桥梁螺旋构象发生改变，转换为弯折状态。转位结束后，桥梁螺旋又恢复到原来的直线构象，与 RNA 链 3′端之间拉开距离，然后重复上述过程。

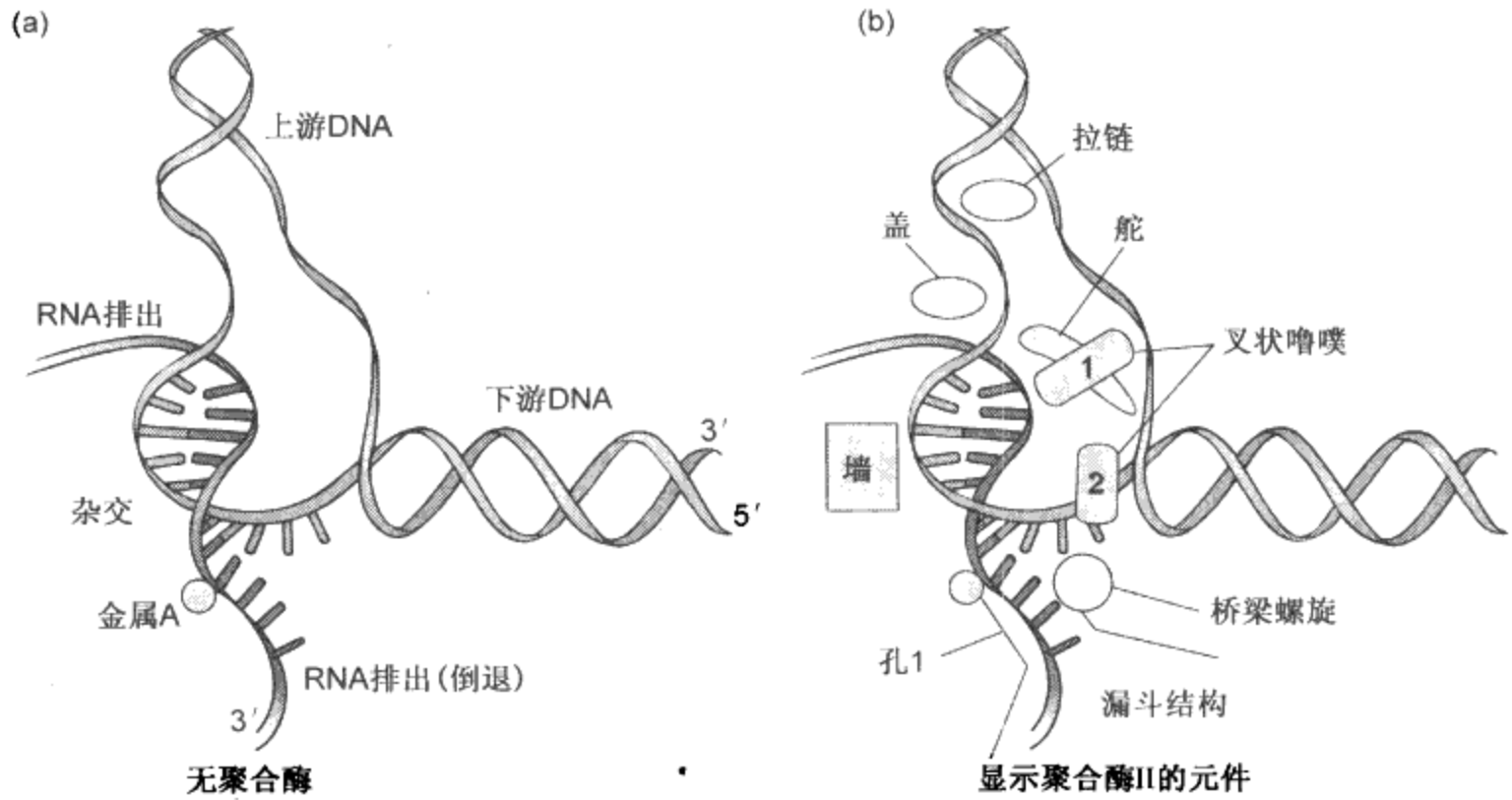


图 10.14　转录泡。(a) 核苷酸的位置。DNA 模板链为蓝色，DNA 非模板链为绿色，RNA 分子为红色，实线表示处于晶体结构中的核苷酸，虚线表示处于晶体结构外的核苷酸。(b) 核苷酸与 RNA 聚合酶 II 中的关键元件。RNA 聚合酶 II 的关键元件叠加在图 (a) 中的核苷酸上，这些元件包括突出于钳的蛋白质环 (拉链、盖、舵)、叉环 1 和 2、桥梁螺旋、漏斗结构、孔 1 和阻隔墙。(*Source*: Adapted from Gnatt, A. L., P. Cramer, J. Fu, D. A. Bushnell, and R. D. Kornberg, Structural basis of transcription: An RNA polymerase II elongation complex at 3.3Å resolution. *Science* 292 (2001) p. 1879, f. 4.)

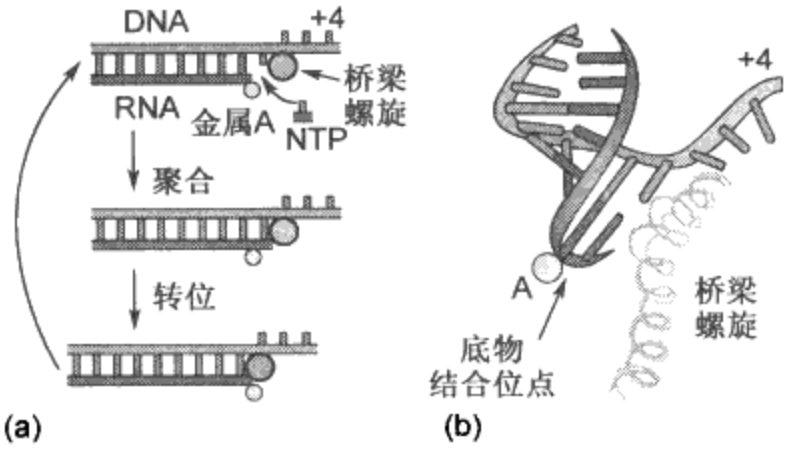


图 10.15　转位机制。(a) 模型。桥梁螺旋最初处于直线构象状态 (橙黄色)，为核苷酸 (NTP) 进入活性位点而空出间隙，活性位点用黄色圆环 (金属 A) 标注。在合成步骤中，核苷酸添加到正在生长的 RNA 链 (红色) 的 3′端，填充 RNA 链的 3′端与处于直线构象状态的桥梁螺旋之间的间隙。在转位步骤中，RNA-DNA 杂交分子向左移动一个碱基的距离，使模板链的下一个核苷酸进入活性位点，同时，桥梁螺旋发生弯曲 (用绿色圆点表示)，继续与 RNA 分子的末端相接触。当桥梁螺旋恢复到直线构象 (左侧箭头) 状态时，再次打开核苷酸入口，使活性位点允许另一个核苷酸的进入。(b) 桥梁螺旋的直线构象和弯曲构象。橙黄色螺旋代表直线构象，绿色螺旋代表弯曲构象。弯曲桥梁螺旋与正在生长的 RNA 链的末端十分靠近。(*Source*: Adapted from Gnatt, A. L., P. Cramer, J. Fu, D. A. Bushnell, and R. D. Kornberg, Structural basis of transcription: An RNA polymerase II elongation complex at 3.3 Å resolution. *Science* 292 (2001) p. 1880, f. 6.)

转位假说的进一步证据来自酵母 RNA 聚合酶 II 与 α-鹅膏蕈碱共结晶体的结构图。α-鹅膏蕈碱的结合位点十分靠近桥梁螺旋，两者之间可形成氢键。α-鹅膏蕈碱的结合会严重抑制转位所必需的桥梁螺旋弯曲构象的形成，这也解释了 α-鹅膏蕈碱可阻断 RNA 的合成但不能阻断核苷酸的进入或磷酸二酯键形成的原因。在磷酸二酯键形成后，α-鹅膏蕈碱可阻断转位，进而抑制 RNA 的合成。

小结　酵母 RNA 聚合酶 II (缺失 Rpb4/7 亚基) 的转录延伸复合体晶体结构图像显示，钳状结构紧紧环抱着位于酶裂隙内的 RNA-DNA 杂交分子，确保转录过程中的进行性。此外，钳的 3 个环结构——舵、盖和拉链分别在下列事件中发挥重要作用：引发 RNA-DNA 杂交体的解离、维持解离状态、维持模板 DNA 的解离。酶的活性中心位于孔 1 的末端，此孔是核苷酸进入酶的通道，也是在聚合酶发生倒退时 RNA 逐出的通道。桥梁螺旋位于活性中心附近，该螺旋的可挠性对转录过程中的转位是至关重要的。α-鹅膏蕈碱与桥梁螺旋附近位点的结合会严重抑制桥梁螺旋弯曲构象的形成，从而阻断了转位。

转位后 RNA 聚合酶 II 的三维结构。Kornberg 及同事在制备转录复合体的晶体时，由于复合体处于进行转录的过程中，其 RNA 的两个末端并不是

处于自然状态。在3′端，聚合酶已发生倒退，没有可接纳下一个核苷酸的空位；在5′端，RNA-DNA杂交分子并没有发生解旋，也就无法鉴定聚合酶在杂交分子解旋过程中的作用。2004年，Kornberg及同事在克服了上述所有的问题后，制备了另一个结晶复合体。研究者将长度为15nt的DNA寡核苷酸链以及长度为9bp且与DNA互补的RNA分子同时与聚合酶结合，形成人工转录复合体，然后通过向反应体系中加入1分子的3′-脱氧腺苷而终止链的延伸，从而使转录复合体在终止转录前已完成了转位。

X射线晶体衍射图像显示，该复合体与以前获得的复合体在核苷酸方面存在着两个显著的差异。第一，RNA链3′端的+1位点处于空位状态，可接纳下一个核苷酸，这正是转位后复合体应有的特征。第二，在RNA链的5′端，RNA链与DNA链在形成8个碱基对的杂交区域后就明显处于解离状态，与其他实验数据相一致，即RNA-DNA杂交分子的长度大约为8bp。

晶体结构还揭示了RNA聚合酶中参与RNA-DNA杂交分子解离的部分。图10.16显示了RNA聚合酶3个环（Rpb1亚基的盖和舵、Rpb2亚基的叉环1）的位置，而以前的晶体结构未能清晰地显示这3个环在聚合酶中的位置。这些蛋白质环所在的位置说明它们在RNA-DNA杂交分子的解离过程中起重要作用。盖可形成导致RNA链与DNA链分离的楔形结构，该楔形结构能够与RNA链的－8位、－9位和－10位碱基发生相互作用，另外，楔形结构中的Phe252可与DNA链的－10位碱基发生相互作用，这些相互作用不仅促进RNA链与DNA链的分离，而且使RNA链与DNA链保持分离状态。

舵似乎并不直接参与链的分离，但可与DNA链的－9和－10位（可能包括－11位）碱基发生相互作用，与盖协同发挥功能，保持链的分离状态。在RNA-DNA杂交分子中，叉环1与RNA链的－5、－6和－7位碱基相互作用，叉环1在该位置不能促进杂交分子的解离，其可能的作用是防止杂交分子过长，使RNA-DNA杂交分子保持约8个碱基对的长度。

小结 X射线晶体衍射图像显示，Rbp1亚基的盖可与RNA-DNA杂交分子中的初生RNA链的碱基相互作用，使杂交分子在8bp区域之后解链。从而导致杂交的碱基对不超过8个，Rbp1亚基的舵可与DNA模板链上的－9和－10位碱基相互作用，与盖协同保持链的分离状态。Rbp2亚基的叉环1可与RNA链的－5、－6和－7位碱基相互作用，使RNA-DNA杂交分子的长度维持在8个碱基对左右。

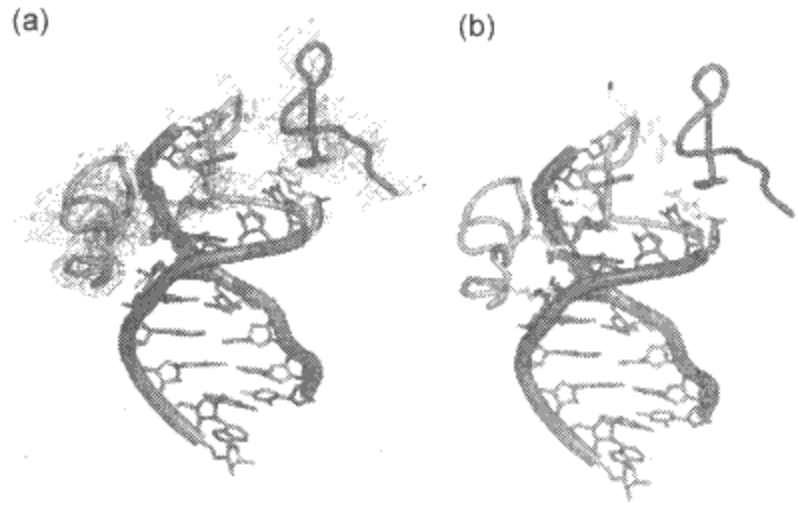


图10.16 Rbp1与Rbp2亚基的3个蛋白质环相对于模板-转录物杂交分子的位置。（a）3个环与杂交分子相互作用的电子密度图示。橙黄色表示Rbp2亚基的叉状环1，绿色表示Rpb1亚基的舵，紫红色表示Rpb1亚基。（b）图中显示了环中与杂交分子发生相互作用的那些氨基酸残基。氨基酸的碳骨架用黄色表示，氮原子用蓝色表示。（*Source*：（a-b）Reprinted with permission from Science, Vol. 303, David A, Bushnell, Kenneth D. Westover, Ralph E. Davis, Roger D. Kornberg, "Structural Basis of Transcription: An RNA Polymerase II-TFIIB Cocrystal at 4.5 Angstroms" Fig. 2 c&d, p. 1015, Copyright 2004, AAAS)

碱基选择的结构原则 在2004年的另一篇论文中，Kornberg及同事报道了不同转位后复合体的X射线衍射晶体衍射图像的分析数据。首先，研究者将一组如图10.17所示的人工合成寡核苷酸与RNA聚合酶II结合，人工合成的寡核苷酸产物包含部分双链DNA模板、长度为10nt的RNA产物，其3′端的核苷酸为3′-脱氧腺苷。我们在前面已了解到，位于末端的3′-脱氧腺苷可阻止其他核苷酸在RNA链上的添加，从而使RNA聚合酶停留在转位后状态。研究者将该复合体的晶体分别浸泡于含有能与DNA模板链的下一个核苷酸正确配对的核苷酸（UTP）的溶液中，或浸泡于含有不能与DNA模板链下一个核苷酸正确配对的核苷酸溶液中。对所得的晶体进行X射线衍射晶体图像分析时发现，

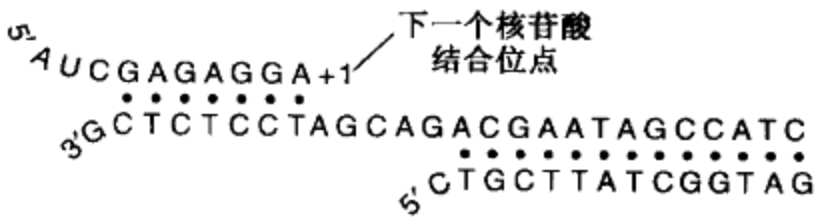


图10.17 用于模拟转录复合体的人工合成的寡聚核苷酸复合物。蓝色表示DNA模板链，绿色表示非模板链，红色表示RNA链，其末端为链终止核苷酸3′-脱氧腺苷（紫色）。

两者之间存在显著的不同：错配核苷酸邻近正确核苷酸应占据的位置，相对于正确核苷酸而言，错配核苷酸发生了倒转（图 10.18）。

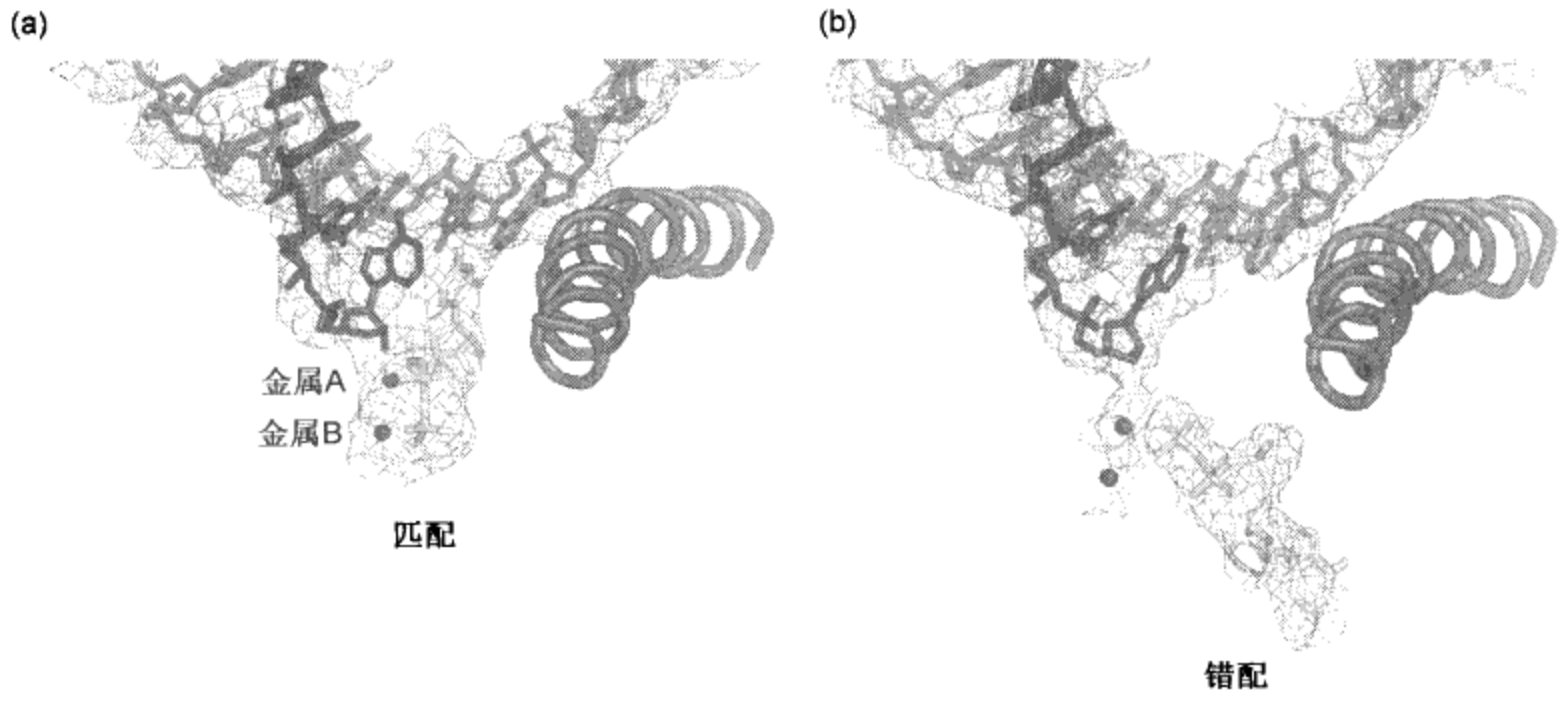


图 10.18　分别位于 A 位点和 E 位点的正确配对（a）与错配（b）的核苷酸。紫红色小球体表示活性位点的金属离子，分别标注为金属 A 和金属 B。DNA 分子为蓝色，RNA 分子为红色，位于 A 位点和 E 位点的核苷酸为黄色。绿色螺旋管为 RNA 聚合酶中的桥梁螺旋。（*Source*: Reprinted from Cell, Vol. 119, Kenneth D. Westover, David A. Bushnell and Roger D. Kornberg, "Structural Basis of Transcription: Nucleotide Selection by Rotation in the RNA Polymerase II Active Center," p. 481-489, Copyright 2004 with permission from Elsevier.）

分析数据显示，在 RNA 聚合酶 II 的活性位点处有两个截然不同的核苷酸结合位点。一个是以前就已知并命名的磷酸二酯键形成位点，即 **A 位点**（A site），第二个是 **E 位点**（E site），即"进入"的意思，核苷酸先与该位点结合，然后才能进入 A 位点。Alexander Goldfard 及同事对 *E. coli* RNA 聚合酶的生化研究已预测了该位点的存在。

这两个位点具有一定的重叠，Kornberg 及同事注意到，核苷酸通过进入通道移向 A 位点时必须先经过 E 位点。从 E 位点进入 A 位点时，核苷酸会发生旋转，这种旋转有何功能性意义？Kornberg 及同事推测可能与核苷酸的专一性有关。正确配对的核苷酸可能比错配的核苷酸更易旋转入位，而且，这种旋转或许会影响核糖核苷酸与模板链上的脱氧核糖核苷酸之间的特异性（核糖的特异性），但是晶体结构图像并没有给出有关这种特异性原则的其他线索（聚合酶的氨基酸侧链并不靠近 A 位点或 E 位点核苷酸的 2′-羟基，不会有氢键的形成，所以只能解释为核糖的特异性）。

晶体结构也进一步证实活性位点确实结合有 2 个金属离子。其中一个离子（金属 A）会永久地结合在酶的活性位点上，而另一个离子（金属 B）进入酶的活性位点后会结合在正被掺入的核苷酸上（结合在 β 磷酸和 γ-磷酸上）。与以前的结构信息相反，最新结构信息显示，这两个金属离子具有相等的结合强度。所以，在 RNA 聚合酶中，磷酸二酯键的形成机制有赖于活性位点上的 2 个金属离子。

小结　在穿过入口移向 RNA 聚合酶 II 活性位点过程中，正被掺入的核苷酸必须先进入 E（进入）位点，而 A 位点是磷酸二酯键形成的位点。相对于在 A 位点时的位置而言，该核苷酸在 E 位点发生了倒转。E 和 A 位点具有一定的重叠。在 E 和 A 位点之间核苷酸发生的旋转可能与碱基和糖的特异性识别有关。活性位点结合有 2 个金属离子（Mg^{2+} 或 Mn^{2+}），其中一个会永久地结合在酶的活性位点上，而另一个进入活性位点，与正被掺入的核苷酸结合。

Rbp4 和 Rbp7 的功能　已讨论过的研究为我们了解 RNA 聚合酶 II 的结构和功能提供了大量信息，但未能提供有关 Rbp4 及 Rbp7 功能的信息，原因是 Kornberg 及同事在获取聚合酶晶体的过程中，Rbp4 和 Rbp7 这两个亚基发生了缺失。为弥补这一缺陷，分别由 Patrick Cramer 和 Kornberg 领导的两个研究小组从酵母中成功获得了完整的 12 亚基聚合酶晶体。为解决 Rbp4 和 Rbp7 亚基缺失的问题，Cramer 的研究小组通过基因克隆策略，在 *E. coli* 表达 Rbp4 和 Rbp7，然后用过量的 Rbp4、Rbp7 与纯化的 10 亚基聚合酶与共同孵育而获得高纯度完整的 12 亚基同

质聚合酶。Kornberg 研究小组在 Rbp4 亚基上添加表位附加，利用表位的特异性抗体进行亲和层析而获得纯化的 12 亚基聚合酶。为进一步增加获得完整全酶的概率，研究者从处于静止期的酵母细胞中提取了 RNA 聚合酶，因为该阶段的酵母细胞内含有 12 亚基全酶的比例要高于 10 亚基的核心酶。

图 10.19 显示的是 Cramer 及同事制备的 12 亚基聚合酶的晶体结构。在该结构图中，因 Rbp4 和 Rbp7 亚基从聚合酶的一侧伸出，从而显现了 Rbp4 和 Rbp7 亚基的结构。这 2 个亚基似楔形，其尖端插入酶的其他部分（即核心酶）内。此外，Cramer 及同事还注意到，Rbp4/7 的存在与否会直接影响钳的位置。Rbp4/7 缺失时，游离的钳状结构处于打开状态，如图 10.19 a 右下方的插图所示；当楔形蛋白 Rbp4/7 存在时，钳状结构在楔形蛋白的束缚下处于闭合状态。

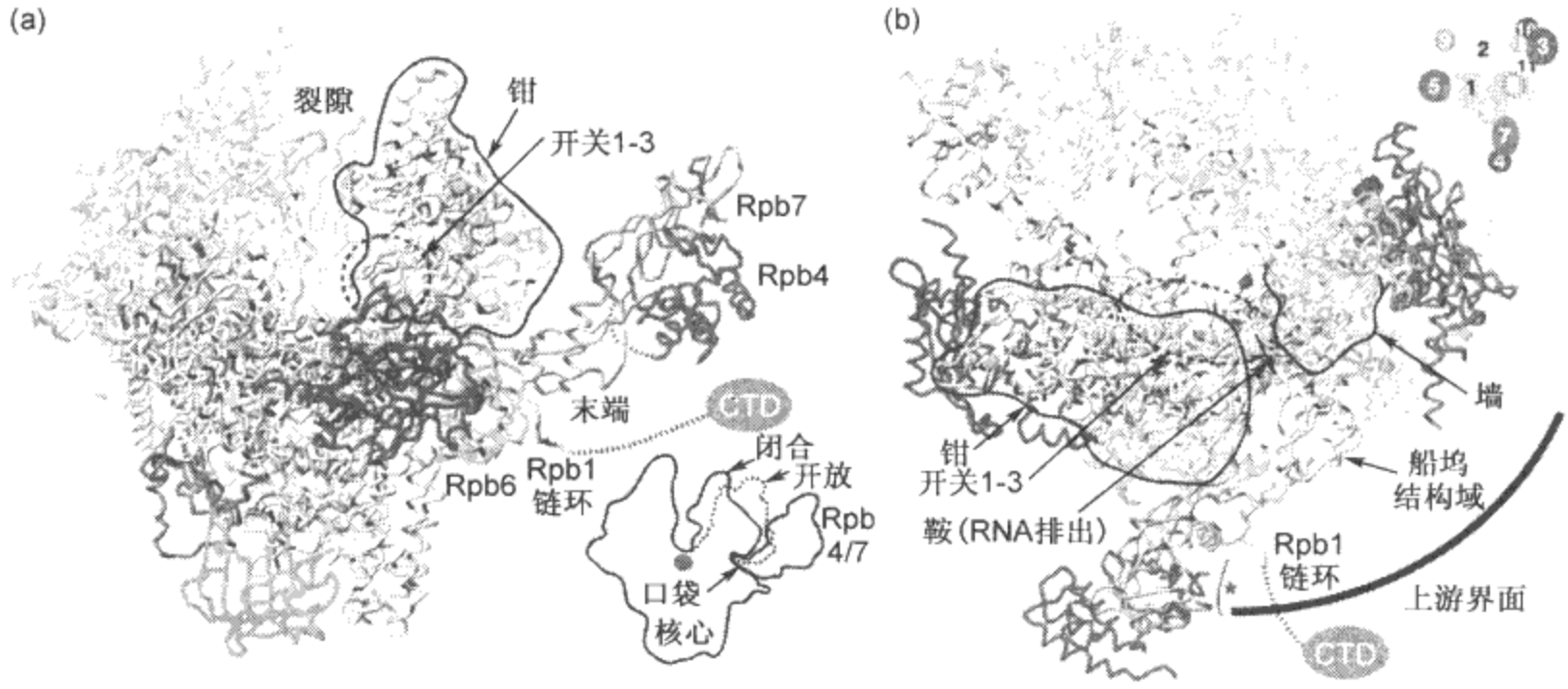


图 10.19　酵母 12 亚基 RNA 聚合酶 II 的晶体结构。（a）Rbp4/7 与核心聚合酶间相互作用的结构图。Rbp4 和 Rbp7 分别用紫红色和蓝色表示，黑色实线勾画出了钳状结构的大体轮廓，开关 1～3 的位置用虚线圆环标注。8 个锌离子用青色球体表示。结合在裂隙（在此图中很难看清）基部活性位点上的镁离子用紫色球体表示，连接在 Rbp1 亚基 CTD 上的链环用虚线表示，右下方的插图显示的是钳状结构处于闭合及打开时的位置。该图表明，Rbp4/7 的结合与钳状结构处于打开状态时的位置是相斥的，即 Rbp4/7 的结合会使钳子闭合。（b）该结构的一个视图，各亚基的颜色如右上插图所示。该视图突出显示了 Rbp4/7 延伸到聚合酶平台结构域后产生的效应。右下方实线圆弧代表以活性位点为中心的 −25bp 片段，这是活性位点与 TATA 盒与转录起始位点之间的最短距离。在图下方中心位置处的蓝色星号表示 Rbp7 亚基上潜在的 RNA 结合位点。（*Source*：(a-b) © 2003 National Academy of Sciences Proceedings of the National Academy of Sciences，Vol. 100，no. 12，June 10，2003，p. 6964-6968 "Architecture of initiation-competent 12-subunit RNA polymerase II," Karim-Jean Armache，Hubert Kettenberger，and Patrick Cramer，Fig. 2，p. 6999.）

在了解聚合酶是如何与启动子 DNA 结合方面，这一新的信息可提供怎样的线索呢？Cramer 及同事以及 Bushnell 和 Kornberg 等认为，当聚合酶核心结合在双链形式的启动子上后，启动子随即发生熔解，Rbp4/7 结合上来并使 DNA 模板链上方的钳闭合，将非模板 DNA 链排除在活性位点以外。但他们也指出，这个简单模型与其他证据相矛盾：首先，从其他生物体中获得的 RNA 聚合酶中也有 Rbp4/7 的同源物，但都不与核心酶发生解离。类似地，*E. coli* RNA 聚合酶全酶的晶体结构呈闭合构象，似乎不允许双链 DNA 结合，而 *E. coli* RNA 聚合酶全酶确实参与了转录的起始（第 6 章）。因此，两组研究者认为，启动子 DNA 结合在酶的外表面并溶解，然后模板链进入酶的活性位点，使启动子 DNA 发生显著弯曲。

两个研究小组都注意到，Rbp4/7 能够对聚合酶与通用转录因子间的相互作用产生强烈的影响，我们将在第 11 章学习相关内容。没有其他几类通用转录因子的协助，RNA 聚合酶 II 自身不能完成与启动子 DNA 的结合，有些通用转录因子可直接与 RNA 聚合酶的"平台区"相互作用。如图 10.19b 所示，Rbp4/7 主要延伸到平台区，所以，Rbp4/7 在结合关键通用转录因子过程中发挥着重要作用。

小结 12 亚基 RNA 聚合酶 II 的结构显示，Rbp4/7 的存在会迫使钳处于闭合状态。由于转录起始由 12 亚基聚合酶引发，其钳状结构处于闭合状态，所以，模板 DNA 链在进入活性位点之前，启动子 DNA 要先行熔解。该结构也显示 Rbp4/7 延伸到聚合酶的平台区，有助于特定通用转录因子的结合，从而促进转录的起始。

10.2 启动子

真核生物有三类结构不同的 RNA 聚合酶，分别转录不同的基因。因此，我们可以预测这三类 RNA 聚合酶能够识别不同的启动子，目前这一推测已经被证实。下面我们将学习 RNA 聚合酶所识别的启动子结构，并对本章的内容做结论性总结。

II 类启动子

因为大家对这类启动子比较熟悉，所以首先介绍 RNA 聚合酶 II 所识别的启动子（**II 类启动子**）结构。这类启动子在某种程度上类似于第 6 章学过的原核生物启动子。II 类启动子由核心启动子（core promoter）和**上游启动子元件**（upstream promoter element）组成。核心启动子可由 4 个元件组成：中心约位于－33 位的 **TATA 盒**（TATA box）、位于 TATA 盒上游的 **TFIIB 识别元件**（TFIIB recognition element，BRE）、以转录起始点为中心的**起始子**（initiator，Inr）、**下游启动子元件**（downstream promoter element，DPE）。除了位置不同之外，核心启动子与上游启动子元件的功能也不相同。在绝大多数的启动子中，由通用转录因子识别核心启动子，帮助 RNA 聚合酶与之结合并形成预引发复合体（preinitiation complex），相关内容将在第 11 章介绍。但对某些少数启动子而言，其上游启动子元件由参与起始过程的其他转录因子识别，相关内容将在本章的后面内容中介绍。图 10.20 总结了所有 5 个启动子元件的相对位置关系。然而，就大多数启动子而言，可能会缺少其中的一个或多个元件。

TATA 盒 到目前为止，对存在于许多 II 类启动子上的保守一致性序列 TATAAA（存在于非模板链上）研究得最为深入。在高等真核生物中，该序列末端的 A 通常位于转录起始位点上游 25～30bp 位置处，该序列因其前 4 个碱基而得名为 TATA 盒。真核生物的 TATA 盒与原核生物的－10 框十分类似，两者最大的区别在于距离转录起始点的位置不同，一个为－25 ～ －30，另一个为－10（酵母 TATA 盒的

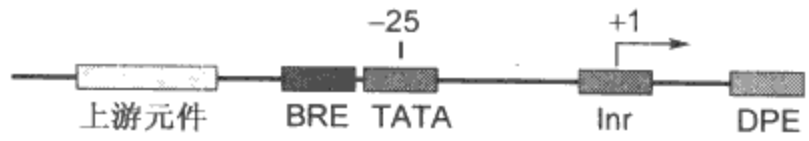


图 10.20 II 类启动子的一般性结构。II 类启动子由核心启动子和上游启动子元件（黄色）组成。核心启动子包含 4 个元件，从 5′端到 3′端依次为：TFIIB 识别元件（BRE，紫色）、TATA 元件（红色）、起始子（Inr，绿色）和下游启动子元件（DPE，蓝色）。通常，II 类启动子即使缺失这些元件中的一个或多个，仍具有正常功能。事实上，DPE 可以取代 TATA 盒，所以这两个元件很少能够同时出现在同一个启动子中。

位置具有可变性，在转录起始点上游的 30～300bp）。

TATA 盒通常包含保守的共有序列，但也有例外，而且这种例外的事例很多。有时在 TATA 盒内会出现 G 和 C，如兔 β-球蛋白基因 TATA 盒的起始序列为 CATA。TATA 盒的共有序列特征通常是不明显的，有两类基因的启动子甚至缺乏 TATA 盒。①一类是**持家基因**（housekeeping gene），这类基因在细胞内都是组成型表达的，其产物控制着基础生化途径，是维持细胞基本生命活动所必需的，如腺嘌呤脱氨酶基因、胸腺嘧啶合成酶基因、二氢叶酸还原酶基因，这些基因编码的酶都是核苷酸合成所必需的。此外，还包括 SV40 编码晚期病毒蛋白的基因。这些基因具有 GC 盒，可补偿 TATA 盒的缺失（第 11 章）。在果蝇（*Drosophila*）中，大约只有 30%的 II 类启动子含有明显的 TATA 盒。但是，许多缺失 TATA 盒的启动子含有 DPE，与 TATA 盒的功能相同的。②另一类缺失 TATA 盒的基因是发育调节基因，如控制果蝇发育的同源异型基因（homeotic gene）、哺乳动物免疫系统发育阶段的活性基因，以及本章后面将要介绍的小鼠末端脱氧核苷酸转移酶 TdT 基因。一般来讲，**特异性基因**（specialized gene）（有时也称为奢侈基因）都具有 TATA 盒，这类基因仅在特定类型细胞内表达，如在皮肤细胞内表达的角蛋白基因、在红细胞内表达的血红蛋白基因。

那么，TATA 盒具有怎样的功能？这可能要视基因而论，探讨这个问题的最初实验是缺失 TATA 盒，然后分析缺失 TATA 盒后启动子的体外转录活性。

1981 年，Christophe Benoist 和 Pierre Chambon 对 SV40 早期启动子进行缺失突变研究。在研究启动子活性的实验中，研究者采用引物延伸法和 S1 作图法（第 5 章）。利用这些技术，我们就可以根据标记 DNA 片段产物的长度和丰度，推定转录的起点

及启动子活性。如图 10.21a 所示，Benoist 和 Chambon 通过逐渐扩大 TATA 盒下游 DNA 序列的缺失长度而获得系列缺失突变体，如 P1A、AS、HSO、HS3 和 HS4 等突变体。缺失的序列中包括转录起始位点，而且每次缺失的碱基对数目恰好等于缩减了 S1 的酶切信号。缺失 S1 酶切信号的结果与因序列缺失导致转录起始位点下移的结果相一致，如果 TATA 盒的作用是将转录起始位点定位在距离 TATA 盒最后一个碱基的下游 25～30bp 处，则起始位点的漂移就是我们预期的结果。如果事实如此，那么将 TATA 盒完全缺失会产生怎样的后果呢？连同 H4 缺失的序列以及 TATA 盒共同缺失的 H2 突变体可以为我们提供答案。图 10.21b 的泳道 8 显示，TATA 盒的缺失导致转录可在多个位点处起始，但不会降低转录效率。如果 TATA 盒影响转录效率的话，那么 S1 信号的增强表明转率效率还增加了。由此可见，TATA 盒的作用显然是参与转录起始位点的定位。

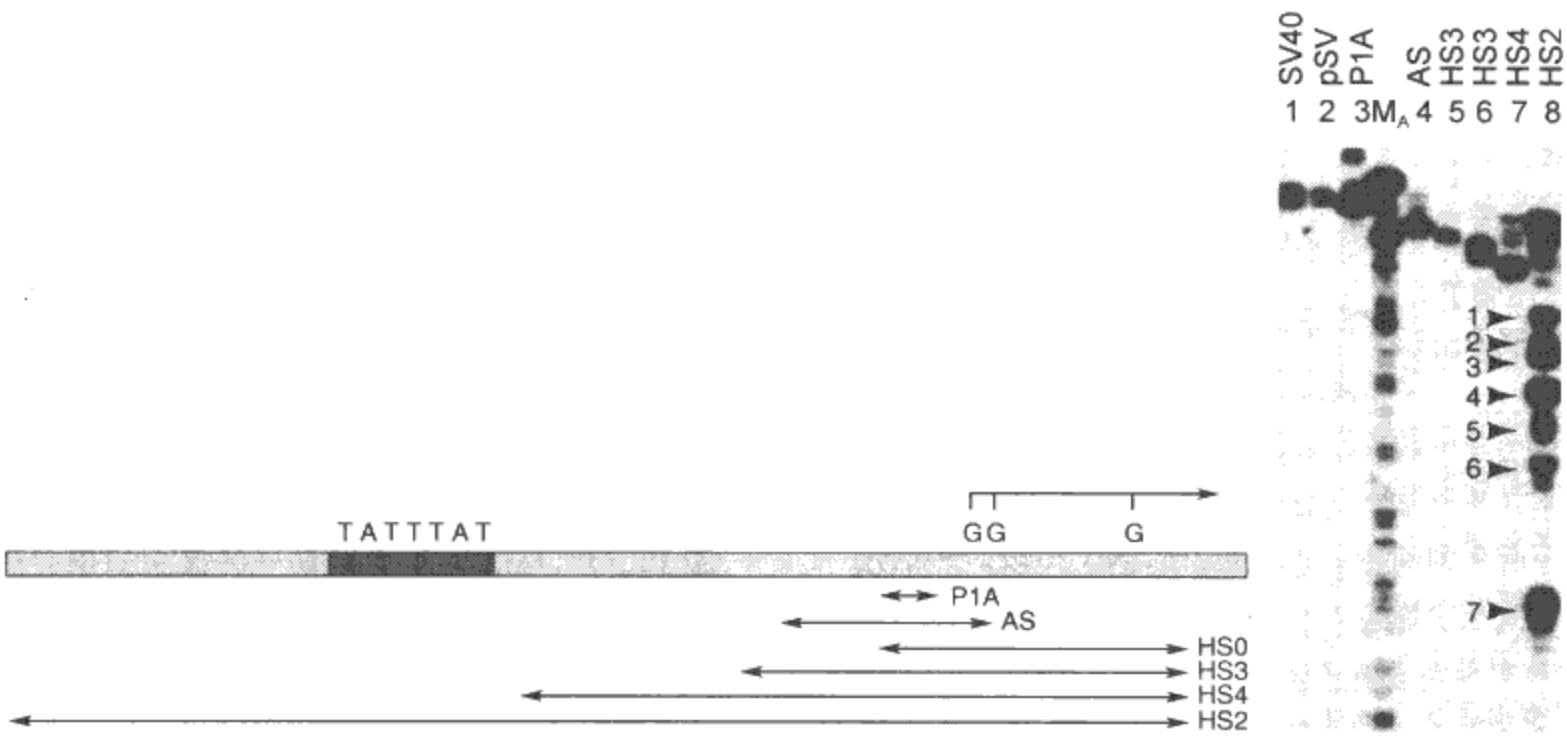


图 10.21　SV40 早期启动子缺失突变的效应。(a) 缺失图谱。突变体的名称标注在箭头的右侧，箭头长度表示每次缺失片段的长度。TATA 盒（TATTTAT，红色）及 3 个转录起始点（均为 G）在图板上方标出。（b）突变体转录起始位点的定位。Benoist 和 Chambon 分别用 SV40 DNA、含有野生型 SV40 早期基因区（pSV1）质粒以及 SV40 早期启动子系列缺失突变后的 pSV1 突变体来转化细胞。研究者用 S1 作图确定转录起始位点。突变体的名称标注在相应泳道的上方。M_A 泳道为分子质量标记，泳道 HS2 中各条带左侧的数字表示异常转录起始位点，这些位点在野生型启动子及其他突变启动子中都未检测到。转录起始位点的异质性显然是由于 TATA 盒缺失造成的。（*Source*：（b）Benoist C. and P. Chambon, In vivo sequence requirements of the SV40 early promoter region. *Nature* 290（26 Mar 1981）p. 306, f. 3.）

通过进一步实验，Benoist 和 Chambon 再次证实上述结论。实验的策略是将 SV40 早期基因转录起始位点与 TATA 盒之间的 DNA 序列进行系统缺失，通过 S1 作图定位各缺失突变体的转录起始位点。野生型基因可在 3 个 G 位点处起始转录，这 3 个 G 簇集在 TATA 盒的下游，与 TATA 盒的第一个 T 相距 27～34bp。当 Benoist 和 Chambon 逐渐扩大转录起始位点与 TATA 盒之间 DNA 序列的缺失长度时，转录不再从这些位点处起始，而是从其他碱基位点起始转录，而且这些碱基通常是嘌呤碱基，这些嘌呤碱基位于 TATA 盒第一个 T 的下游约 30bp 处。也就是说，TATA 盒与转录起始位点间的距离是固定的，并不要求十分精确的起始位点序列。

在这个实验事例中，TATA 盒显然对于转录起始位点的定位是十分重要的，但不会对转录起始效率产生调节作用。然而，有些启动子缺失 TATA 盒后会削弱其功能，如转录效率下降以至检测不到、从异常位点起始转录等。

Steven McKnight 和 Robert Kingsbury 对疱疹病毒的胸苷激酶（tk）基因启动子的研究，为我们提供了另一个有关 TATA 盒功能的事例。研究者采用了**接头分区突变**（linker scanning mutagenesis）的方法，如图 10.22 所示，用长度为 10bp 的接头对 *tk* 基因的启动子序列进行系统替换。接头扫描突变实

验的结果之一是，TATA 盒内序列的突变会破坏启动子的活性（图 10.23），在启动子活性最低的突变体（LS-29/-18）中，TATA 盒区域的正常 GCATATTA 序列突变为 CCGGATCC。

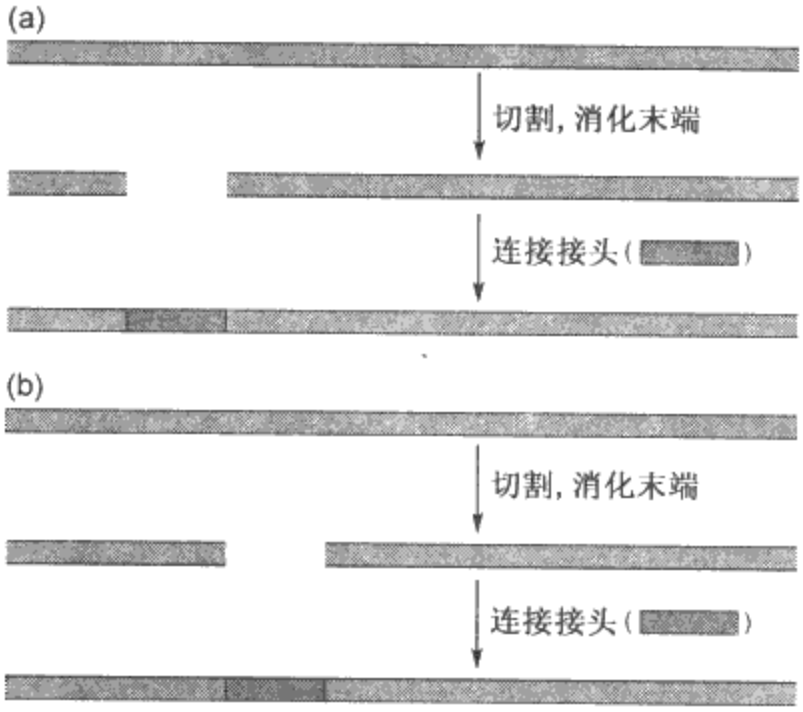


图 10.22　接头分区机制。（a）和（b）是简化程序。用限制性内切核酸酶切割 DNA，并消化末端序列使之从切点处缺失大约 10bp 的序列，即图中所示的间隙，产生两个 DNA 片段。然后用合成的一小段 DNA 接头（红色）填补间隙，将 DNA 片段连接起来。图（b）中的间隙与图（a）中的间隙相比发生了右移，所以图（b）中的接头取代的 DNA 也发生了右移。如此重复，间隙不断右移，就可以用 10bp 的替代整个 DNA。

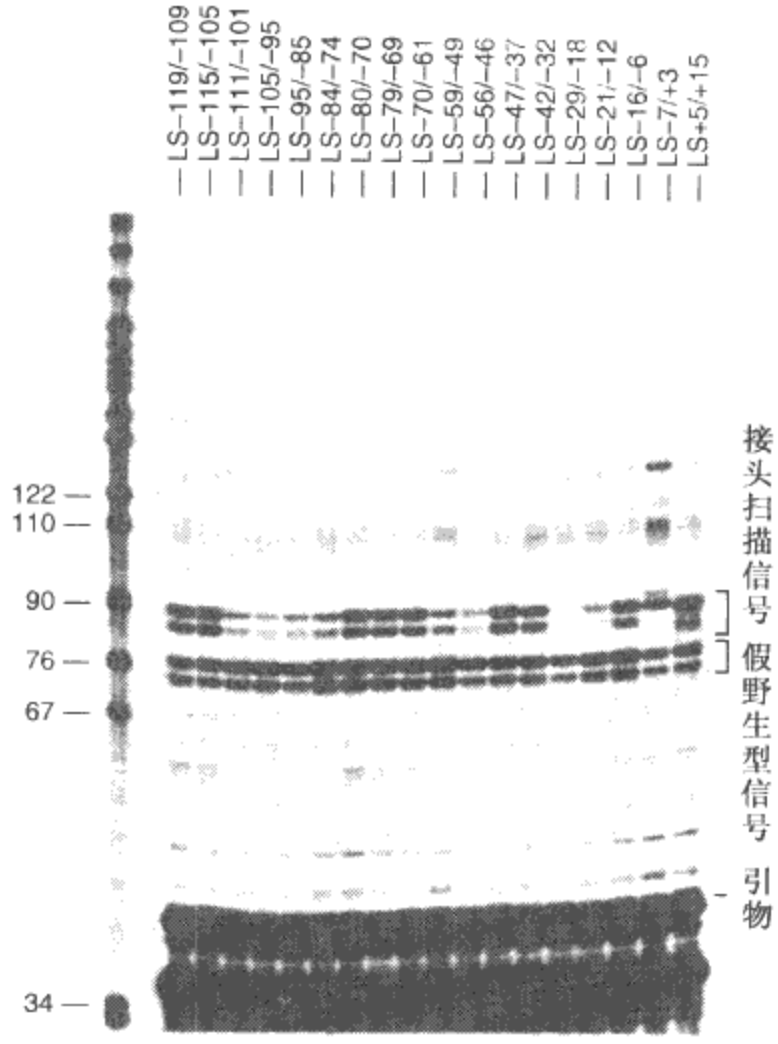


图 10.23　疱疹病毒 *tk* 启动子的接头扫描突变效应。McKnight 和 Kingsbury 对 *tk* 启动子全长序列进行了接头分区突变，然后将突变 DNA 分子注入蛙卵母细胞，同时也注入类似于野生型的 DNA 分子（在＋21 至＋31 处发生突变），这种假野生型突变启动子与正常野生型启动子一样，具有正常的起始转录活性，可以作为内参。研究者通过引物延伸法分析实验质粒与对照质粒的转录情况，如所预期，对照质粒的转录相对稳定，而实验质粒的转录水平具有很大的变化，这要取决于突变发生的位点。（*Source*：Adapted from McKnight, S. L. and R. Kingsbury, Transcriptional control signals of a eukaryotic protein-coding gene. *Science* 217 (23 July 1982) p. 322, f. 5.）

由此可见，有些 II 类启动子需要 TATA 盒发挥功能，但有些 II 类启动子仅需要 TATA 盒来定位转录起始点。此外，有些 II 类启动子特别是持家基因的启动子根本就没有 TATA 盒，但仍然具有正常的功能。如何解释启动子间的这些差异呢？正如我们将在第 11 章和第 12 章学习到的那样，启动子起始转录的活性有赖于转录因子和 RNA 聚合酶组装形成的**前转录起始复合体**（preinitiation complex），该复合体在转录起始位点形成并启动转录过程。在 II 类启动子中，TATA 盒是蛋白质转录因子组装的场所。第一个与 TATA 盒序列结合的蛋白质是 TFIID，其中包括 TBP，具有招募其他因子的作用。即使启动子中缺失 TATA 盒，前起始复合体的组装仍需要 TBP，但因为没有 TATA 盒供其结合，TBP 可借助于结合在其他启动子元件上的蛋白质而将自己定位在正确的位置上。

起始子、下游启动子元件和 TFIIB 识别元件

有些 II 类启动子在转录起始位点附近具有保守性序列，是起始最佳转录所必需的。这些序列称为起始子（initiator），哺乳动物起始子的共有序列为 PyPyAN（T/A）PyPy，其中 Py 代表嘧啶碱基（C 或 T），N 代表任一碱基，带下画线的 A 是转录起始点。果蝇起始子的保守序列为 TCA（G/T）T（T/C）。有关起始子的经典事例来自腺病毒的主要晚期启动子，其起始子与 TATA 盒共同组成核心启动子，能够启动下游任一基因的转录，只是效率较低。该启动子能够被上游元件或邻近的增强子所激活。

另一个带有重要起始子的基因是哺乳动物的末端脱氧核苷酸转移酶（TdT）基因，该基因在 B 和 T 淋巴细胞发育过程中被激活。Stephen Smale 和 David Baltimore 在研究小鼠 TdT 启动子时发现，该

启动子无 TATA 盒及明显的上游启动子元件，但是含有起始子序列。该起始子足以有效地起始 *TdT* 基因本底水平上的转录，此转录是从起始子内部的单一位点处起始的。他们还发现，SV40 启动子的 TATA 盒或 GC 盒能极大地促进 *TdT* 基因从起始子处起始转录。因此，仅由起始子组成的功能性简单启动子其转录效率可以被其他启动子元件所增强。

在果蝇基因的启动子中，普遍存在下游启动子元件（down stream promoter element，DPE）。事实上，在 2000 年 Alan Kutach 和 James Kadonaga 就报道了一个惊人的发现，果蝇基因启动子中 DPE 就像 TATA 盒一样普遍存在。这些 DPE 位于转录起始点下游约 30bp 处，其保守一致性序列为 G（A/T）CG。可以补偿因 TATA 盒缺失造成的转录抑制。事实确实如此，许多正常缺失 TATA 盒的果蝇启动子都包含 DPE，这可解释该种生物富含 DPE 的原因。在缺失 TATA 盒的果蝇启动子中，DPE 与 Inr 是成对存在的，TATA 盒与 DPE 的功能相似性在于它们都能够与关键的通用转录因子TFIID（第 11 章）相结合。

另一个重要的通用转录因子是 TFIIB，与 TFIID、RNA 聚合酶 II 以及其他因子结合在启动子处，组装成前起始复合体而起始转录。有些启动子紧邻在 TATA 盒上游存在着能够帮助 TFIIB 与 DNA 结合的元件，这些 DNA 元件称为 **TFIIB 识别元件（BRE）**，其保守一致性序列为（G/C）（G/C）（G/A）CGCC。

小结 II 类启动子可包含多达 5 个保守元件：由 4 个元件构成的核心启动子及上游启动子元件。4 个核心启动子元件包括 TFIIB 识别元件（BRE）、TATA 盒、起始子（Inr）和下游启动子元件（DPE）。在大多数启动子中，至少会缺失其中的一个元件。事实上，缺失 TATA 盒的启动子通常都含有 DPE，至少在果蝇中情况是这样的。高效表达特异基因的启动子倾向于含有 TATA 盒，而起始持家基因表达的启动子一般缺失 TATA 盒。包括哺乳动物 TdT 启动子在内的某些启动子只有起始子序列，可能还会带有 DPE。

上游元件 McKnight 和 Kingsbury 对疱疹病毒 *tk* 基因的接头分区分析还发现了除 TATA 盒之外的其他重要启动子元件。图 10.23 显示，当－47～－61和－80～－105 这两个区域发生突变时，会引起启动子活性的显著下降。在非模板链上，这两个区域的序列分别为 GGGCGG 和 CCGCCC，即所谓的 **GC 盒**（GC box），存在于各类启动子中，且通常位于 TATA 盒的上游，值得关注的是，在疱疹病毒的 *tk* 启动子中，两个 GC 盒在其两个位点中的方向是相反的。

Chambon 及同事发现，在 SV40 早期启动子内也存在 GC 盒，其拷贝数为 6 个，而不是 2 个，这些元件发生突变会显著降低启动子的活性（表 10.4）。例如，一个 GC 盒的缺失会使转录水平下降到野生型的 66%，如果再缺失一个 GC 盒，则转录水平一直下降到对照水平的 13%。在第 12 章我们将了解到，特异性转录因子 Sp1 与 GC 盒结合后能够增强转录效率，在本章的后面内容中，将介绍能够增强转录效率的 DNA 元件——增强子。与启动子相比，增强子具有 2 个重要的不同，即增强子无位置和方向依赖性，GC 盒虽然无方向依赖性，在倒转 180°后仍具有正常功能（疱疹病毒的 *tk* 启动子会自然发生倒转现象），但与典型的增强子不同，GC 盒具有位置依赖性，而增强子可以在远离启动子数千碱基对的较远位置处发挥作用，甚至位于基因编码区的下游也能发挥增强转录的功能。而 GC 盒如果被移至距离 TATA 盒几十个碱基对的远处，就会失去促进转录的能力。因此，将 GC 作为启动子元件比作为增强子更为恰当，至少对疱疹病毒的 *tk* 基因和 SV40 早期基因而言是这样的。此外，增强子与 GC 盒的差别是细微的，且语义上也模棱两可。

表 10.4　GC 盒突变对 SV40 早期启动子起始基因表达的影响

完整 GC 盒的数目	表达水平/%
6	100
5	66
4	13
3	30
2	40
1	9
0	13

在各种 II 类启动子中还存在一种称为 **CCAAT 盒**（读为“cat”盒）的上游元件。事实上，疱疹病毒 *tk* 启动子就含有 CCAAT 盒。在前面讨论的接头扫描分析中，当 CCAAT 盒发生突变时，并没有检测到启动子活性的下降。但其他研究清晰地显示，CCAAT 盒在疱疹病毒 *tk* 启动子及其他许多启动子中具有重要的作用。同 GC 盒具有自己的转录因子一样，CCAAT 盒也必须与转录因子［**CCAAT 结合转录因子**（CCAAT-binding transcription factor，**CTF**）］结合才能发挥促进转录的效应。

小结 上游启动子元件通常位于 II 类核心启动子的上游，与核心启动子不同的是，上游启动子元件通常与相应的基因特异性转录因子结合。例如，GC 盒与转录因子 Sp1 结合，而 CCAAT 盒与转录因子 CTF 结合。与核心启动子不同，上游启动子元件无方向依赖性，但与典型的增强子相比，上游启动子元件具有位置依赖性。

I 类启动子

RNA 聚合酶 I 识别的启动子又是怎样一种情况？我们之所以将 I 类启动子看做是单一类型的启动子，是因为几乎所有物种只有一类可被 RNA 聚合酶 I 识别的基因，即 rRNA 前体的基因。只有锥虫（trypanosome）例外，除了转录 rRNA 前体基因外，RNA 聚合酶 I 还转录另两类蛋白质编码基因。rRNA 前体基因在每个细胞内都有数百个拷贝，而且这些拷贝都是完全相同的，都具有相同的启动子序列。然而，这一序列因物种的不同而表现出很大的变化，比 RNA 聚合酶 II 识别的启动子序列具有更大的可变性，II 类启动子一般都具有共同的保守序列，如 TATA 盒。I 类启动子的保守序列是富含 AT 的起始子（rINR），位于转录起始位点附近。

Robert Tjian 及同事用接头分区突变分析鉴定出了人类 rRNA 启动子的重要区域，图 10.24 显示了接头分区突变分析的结果，该启动子有两个关键区域，如果这两个区域发生突变会极大地降低启动子起始转录的强度。其中一个是核心元件（core element），位于转录起始点附近，从－45 位延伸到＋20 位；另一个序列为**上游启动子元件**（upstream promoter element，UPE），位于－156 位与－107 位之间。

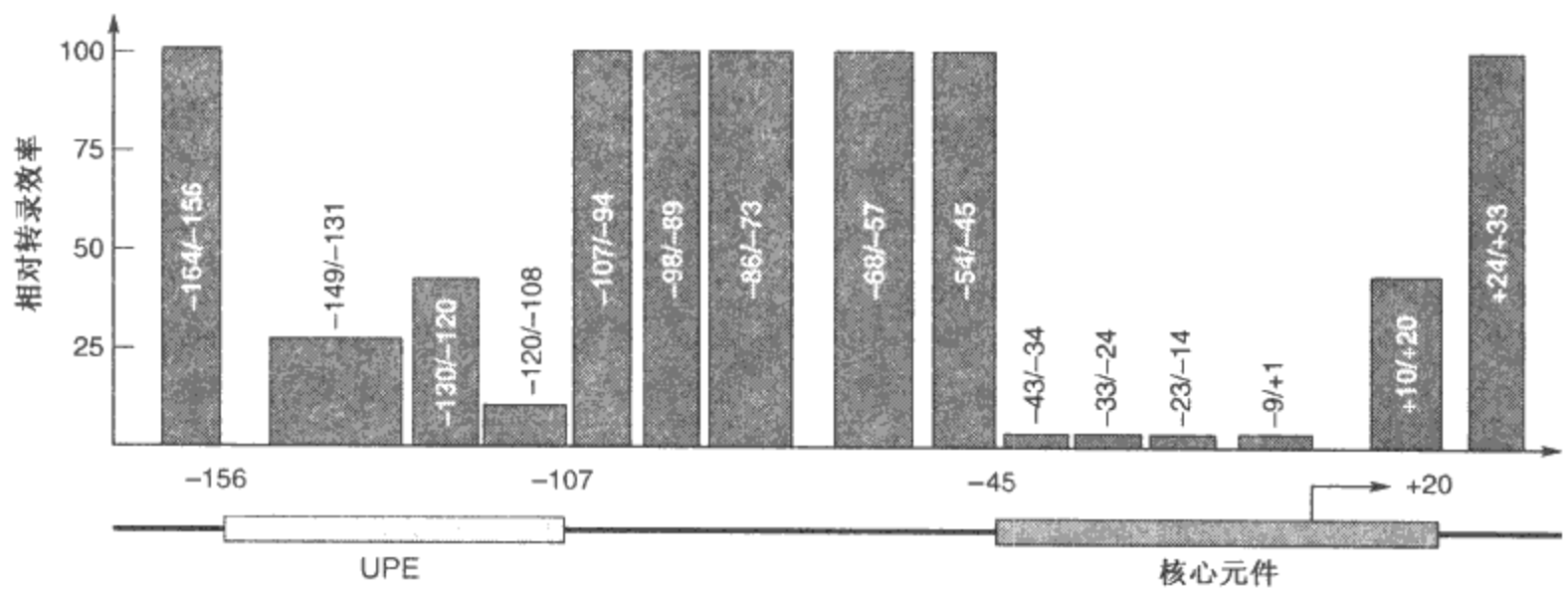


图 10.24　rRNA 启动子的两个元件。Tjian 及同事用接头分区突变法对人类 rRNA 基因的 5′-侧翼区进行系统突变，然后利用体外转录体系检测各个区段突变后对启动子起始转录活性的影响。研究者用柱形图来说明实验结果，结果分析显示，启动子有两个重要的区域：UPE（上游启动子元件）和核心元件。UPE 是最佳转录所必需的，但本底水平的转录不一定需要 UPE 的参与。另一方面，核心元件是任何转录所必需的。（*Source*：Adapted from Learned，R. M.，T. K. Learned，M. M. Haltiner，and R. T. Tjian，Human rRNA transcription is modulated by the coordinated binding of two factors to an upstream control element. *Cell* 45：848，1986.）

两个启动子元件的存在产生了两者间隔距离的重要性问题。本例中，两个元件的间距十分重要，Tjian 及同事在人类 rRNA 启动子的 UPE 与核心元件之间插入或缺失不同长度的 DNA 序列，检测空间距离的改变对启动子起始转录活性的影响。当在启动子元件间删除长度为 16bp 的片段后，启动子强度下降到野生型的 40%，当删除长度为 44bp 的片段后，启动子强度只有野生型的 10%。此外，在启动子元件间添加长度为 28bp 的片段后，不会影响启动子强度，但添加片段长度达 49bp 时，启动子强度降低 70%。由此可见，序列的删除比插入更易影响启动子起始转录的效率。

小结　在不同的物种间，I 类启动子序列的保守性不高，但在一般结构上还是具有较高的保守性。I 类启动子由两个元件构成：位于转录起始位点附近的核心元件和大约在上游 100bp 远处的上游启动子元件（UPE）。这两个元件的空间距离是十分重要的。

III 类启动子

正如我们所了解的那样，RNA 聚合酶 III 可转录多种编码小 RNA 的基因，这些基因包括：①“典型”III 类基因，包括 5S rRNA 基因、tRNA 基因、腺病毒 VA RNA 基因；②一些最近才发现的 III 类基因，包括 U6 snRNA 基因、7SL RNA 基因、7SK RNA 基因以及 EB 病毒的 EBER2 基因。“非典型”III 类基因

的启动子类似于 II 基因的启动子，相反，“典型” III 类基因的启动子完全位于基因的内部。

具有内部启动子的 III 类基因 Donald Brown 及同事首次对爪蟾（*Xenopus borealis*）5S rRNA 基因的 III 类启动子进行了分析，其研究结果令人惊讶，RNA 聚合酶 I 和 II 以及细菌 RNA 聚合酶所识别的启动子大部分都位于基因的 5′-侧翼区，而 5S rRNA 启动子却位于其控制基因的内部。

5S rRNA 启动子位于基因内部的结论是通过以下实验操作得出的：首先，确定启动子的 5′端。Brown 及同事对爪蟾 5S rRNA 基因的 5′端序列进行逐渐缺失，每次缺失序列的长度都会增长，由此获得 5′端长度不等的系列 5S rRNA 基因突变体，体外观察突变对基因转录的影响。通过凝胶电泳测定转录产物的大小，依此判断转录的正确与否。长度约为 120bp（5S rRNA 的大小）的 RNA 是基因的正常转录产物，即使该转录物的序列与真正 5S rRNA 不同。研究者之所以转录获得了非正确序列的转录本，是因为他们因破坏启动子而改变了基因内部的序列。

实验结果让人意外，将基因 5′侧翼区全部删除对转录没有多大的影响（图 10.25）。而且，将基因自身 5′端的较大片段删除后，仍能产生长度约为 120nt 的转录物。然而，当删除的片段超过＋50 位时就会破坏启动子的功能。

利用类似方法，Brown 及同事在基因的转录序列内鉴定出一个位于第 50 位碱基至第 83 位碱基的敏感区域，只要其中的序列不改变，启动子的功能就不会遭到破坏。很明显，这两个位点就是爪蟾 5S rRNA 基因内部启动子的外边界。其他一些实验的研究结果表明，在该区域的外侧插入较大的 DNA 片段，不会对启动子造成损害。Roeder 及同事对启动子内的全部碱基进行系统突变分析，从中鉴定出 3 个区域，只要这 3 个区域不发生改变，启动子功能就不会大幅度降低。这些敏感区称为 A 盒、中间元件（intermediate element）以及 C 盒（无 B 盒，因为在其他的 III 类基因内已发现了 B 盒，但在 5S rRNA 启动子内还没有发现与之相对应的序列）。图 10.26 总结了对 5S rRNA 启动子进行实验分析获得的结果。对另两个典型的 III 类基因，即 tRNA 基因和 VA RNA 基因进行了相同的实验，结果表明这两个基因的启动子都包含有 A 盒与 B 盒（图 10.26b）。其中 A 盒序列与 5S rRNA 启动子的 A 盒序列相似。在一定程度上改变这两个区域的空间距离不会破坏启动子的功能。然而，但这种改变是有限度的，如果在 A 盒与 B 盒之间插入太大的 DNA 片段，则会降低启动子的活性。

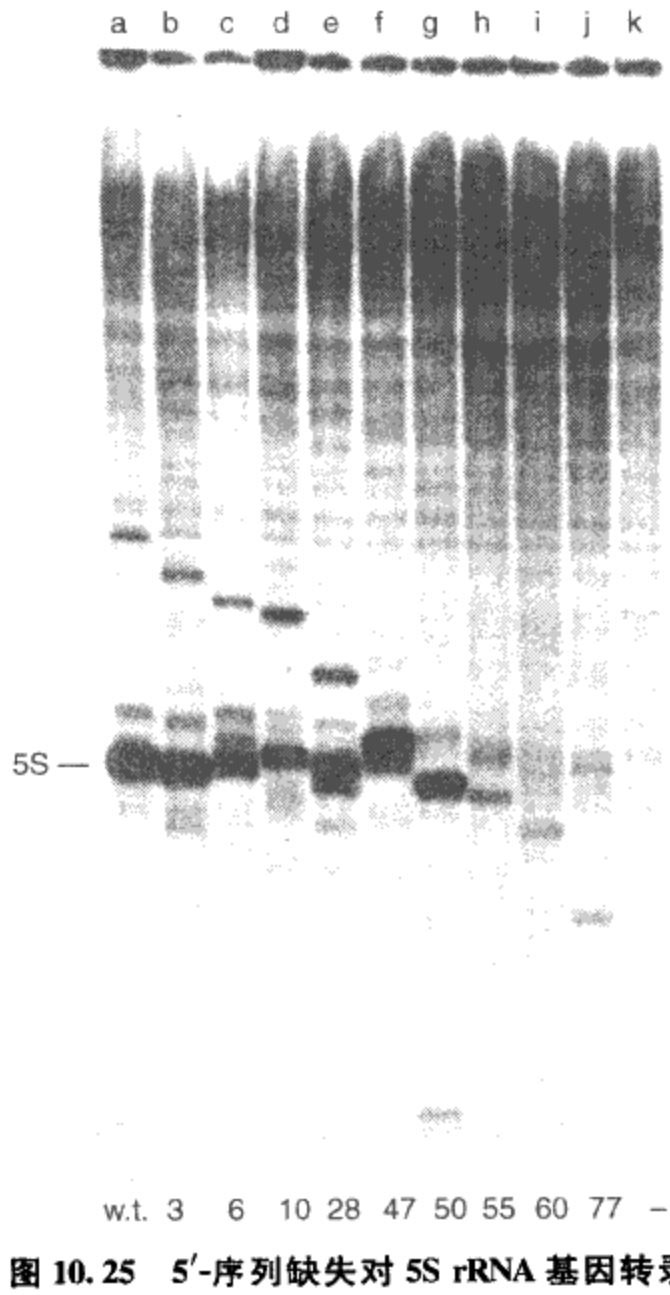


图 10.25 5′-序列缺失对 5S rRNA 基因转录的影响。Brown 及同事对爪蟾 5S rRNA 基因的 5′端序列进行逐渐缺失，然后在体外存在标记底物的条件下转录这些突变基因，对转录产物进行凝胶电泳。实验所用的 DNA 模板分别是：泳道 a，未发生序列缺失的阳性对照；泳道 b～j，序列缺失后的基因，其 5′端碱基位置在图下方标注（如泳道 b，5S rRNA 序列缺失后，突变基因的 5′端相当于野生型基因原来的＋3 位）；泳道 k，阴性对照（pBR322 DNA，无 5S rRNA 基因序列）。泳道 a～g 的系列缺失突变模板具有较强的全长的 5S rRNA 合成能力，其中泳道 g 的模板缺失的序列达到＋50位。若缺失序列的长度继续延长，超过＋50 位时，转录停止。泳道 h～k 也在正常位置处出现了电泳条带，但与 5S rRNA 基因的转录无关，可能是不相干的杂带。（*Source*：Sakonju，S.，D. F. Bogenhagen，and D. D. Brown. A control region in the center of the 5S RNA gene directs specific initiation of transcription：I. The 5′border of the region. *Cell* 19（Jan 1980）p. 17，f. 4.）

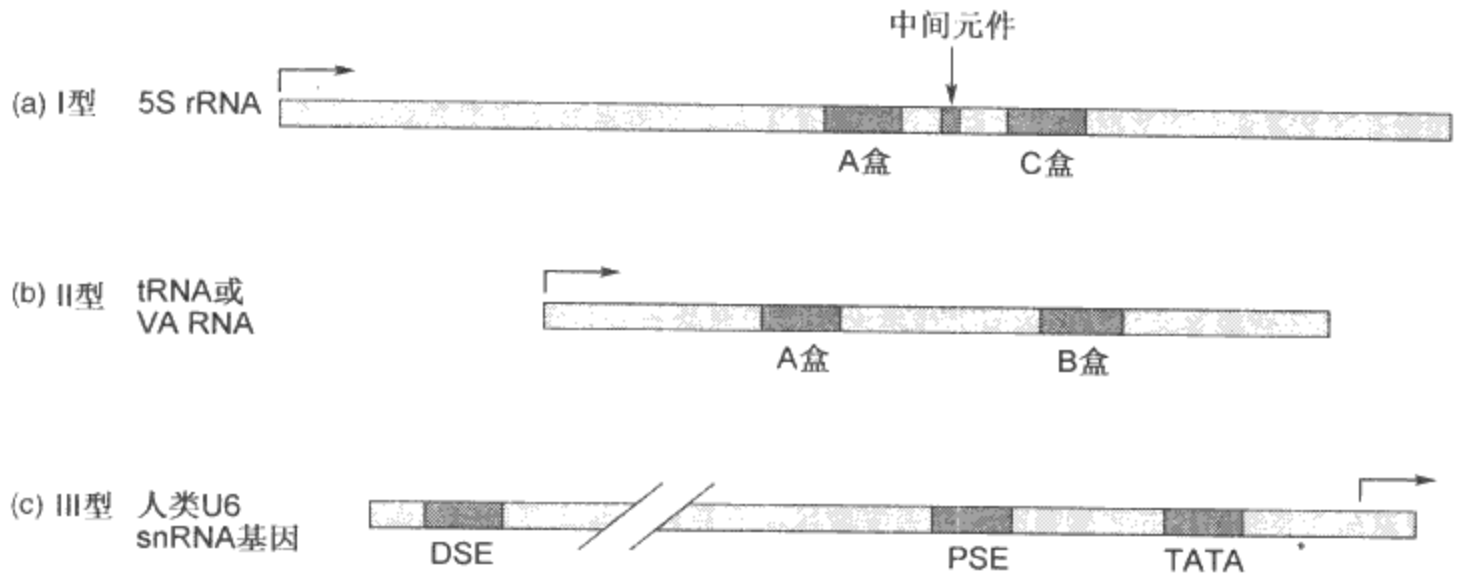


图 10.26 某些 RNA 聚合酶 III 所识别基因的启动子。5S rRNA 基因启、tRNA 基因和 U6 RNA 基因启动子内的基因调控元件以蓝色条框显示。

因此，III 类启动子具有几种不同的类型。5S rRNA 基因的启动子单独划分为一组，属于 I 型启动子（图 10.26a），注意，在本节我们只讨论 III 类启动子，不要将 I 型启动子与 I 类（class I）基因相混淆。第二组为 II 型（type II）启动子，绝大多数的 III 类启动子都属于 II 型启动子，类似图 10.26b 所示的 tRNA 启动子和 VA RNA 启动子。第三组为 III 型（type III）启动子，包括基因的调控元件限定在 5′-侧翼区的非典型启动子，人类 7SK RNA 基因启动子和 U6 RNA 基因启动子是 III 型启动子的典型代表（图 10.26c）。顺便提一下，U6 RNA 是核内小 RNA（snRNA）的成员，在 mRNA 剪接过程中发挥关键作用，相关内容将在第 14 章讨论。最后，还有些启动子（如 7SLRNA 基因的启动子）属于 II 型与 III 型之间的杂合类型，即具有 II 型启动子和 III 型启动子的特点，在基因内部和外部均存在对启动子转录活性具有重要作用的元件。

小结 RNA 聚合酶 III 转录一套较短的基因，典型 III 类基因（I 型和 II 型）启动子的全部序列均位于基因内部。III 类基因（5S rRNA 基因）的 I 型内部启动子可分为 3 个区域：A 盒、短的中间元件、C 盒。III 类基（如 tRNA 基因）的 II 型内部启动子可分为 A 盒与 B 盒 2 个区域。非典型的（III 型）III 类基因的启动子类似于 II 类基因的启动子。

具有类似于聚合酶 II 启动子的 III 类基因 在 Brown 等提出 III 类基因内部启动子的新概念后，人们普遍认为所有 III 类基因都是以这种方式起始转录的。然而，在 20 世纪 80 年代中期发现了例外。1985 年，Elisabetta Ullu 和 Alan Weiner 对野生型和突变型 7SL RNA 基因进行体外转录研究时发现，基因的 5′-侧翼区是基因高水平转录所必需的，如果该区域缺失，转录的效率就会降低为原先的1/100～1/50。Ullu 和 Weiner 推断，对 7SL RNA 基因转录最重要的 DNA 元件位于基因的上游。然而，5′-侧翼区缺失后，突变基因仍然能够转录的事实表明，基因还含有较弱的内部启动子。这些数据有助于解释人类基因组中存在数百拷贝的 7SL RNA 假基因（7SL 基因的无功能拷贝）以及与 *Alu* 基因相关的序列在体内弱转录的原因，这些假基因缺少高水平转录所必需的上游元件。

Marialuisa Melli 及同事注意到，7SK RNA 基因没有类似于典型 III 类启动子那样的内部序列，其 5′-侧翼区与 7SL RNA 基因的 5′-侧翼区同源，基于这些观察，研究者认为 7SK RNA 基因具有一个完整的外部启动子，为证实这一推论，他们连续渐进地对该基因的 5′-侧翼区进行缺失突变，然后测试缺失片段对基因体外转录的效应。图 10.27 显示，删除－37 位上游的序列时，该基因仍能高效表达，合成大量的 7SK RNA。但删除－37 位下游的序列时，转录就不能发生了。此外，相对而言，编码区并非转录所必需，体外转录分析实验中使用的一组突变体，即使缺失发生在编码区内甚至编码区完全缺失时，转录仍能发生。由此可见，7SK RNA 基因没有内部启动子。

位于转录起始点上游约 37 个碱基距离处的启动子有怎样的性质呢？令人感兴趣的是，在这个区域有 TATA 盒序列，改变其中的 3 个碱基（TAT→GCG）会导致转录效率下降 97%。因此，TATA 盒是启动子充分发挥功能所必需的。这一切也许会让你感到困惑，难道 7SK RNA 基因由 RNA 聚合酶 II 而不是 RNA 聚合酶 III 负责转录。如果该基因确实由 RNA 聚合酶 II 负责转录，那么低浓度的 α-鹅膏蕈碱就能抑制转录的发生，但实际上只有较高浓度的 α-鹅膏蕈碱才能抑制 7SK RNA 的合成。从 α-鹅膏蕈碱抑制 7SK RNA 合成的曲线可以看出，我们的预测是正确的，7SK RNA 基因由 RNA 聚合酶 III

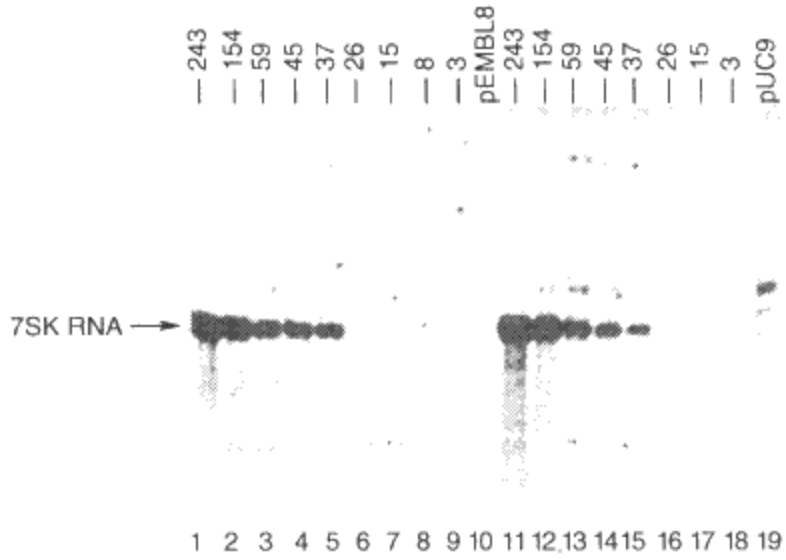


图 10.27　5′端序列缺失突变对 7SK RNA 启动子的影响。Melli 及同事将人 7SK RNA 基因的 5′-侧翼区序列缺失，然后体外转录这些突变基因，凝胶电泳检测 7SK RNA 是否能够正常合成。泳道上方的负数表示突变基因其 5′端保留序列的长度。例如，泳道 9 所用的 DNA 模板其 5′-侧翼只有 3 个碱基，即标注为－3。泳道 1～10：突变基因克隆到 pEMBL8 载体；泳道 11～19：突变基因克隆到 pUC9 载体；泳道 10 和 19 分别表示克隆载体自身转录。对比泳道 5 和 6（或 15 和 16）可以看出，5′-侧翼的第－37 至－26 之间的序列缺失会导致启动子活性的急剧下降，表明这 11 个碱基序列是启动子的一个重要元件。（*Source*：Murphy，S.，C. Diliegro，and M. Melli，The in vitro transcription of the 7SK RNA gene by RNA polymerase III is dependent only on the presence of an upstream promoter. *Cell* 51（9）（1987）p. 82，f. 1b.）

而不是 RNA 聚合酶 II 负责转录。

现在我们已了解到，U6 RNA 基因和 EBER2 基因也属于这种非典型的 III 类基因，具有类似于 RNA 聚合酶 II 所识别的启动子，但却由 RNA 聚合酶 III 负责转录。到第 11 章我们就对这种现象不会感到奇怪了，因为 TATA 结合蛋白（TBP）除了参与 II 基因的转录外，还参与 III 类（和 I 类）基因的转录。

小结　至少有一种 III 类基因，即 7SL RNA 基因具有一个较弱的内部启动子，同时，该基因还拥有高效转录所必需的 5′-侧翼区。其他非典型的 III 类基因（如 7SK 基因和 U6 RNA 基因）无内部启动子，但在基因 5′-侧翼区具有与 II 类启动子十分相似的启动子区域，此外，还包含 TATA 盒。

10.3　增强子与沉默子

许多真核基因特别是 II 类基因都有顺式作用（*cis*-acting）DNA 元件。从严格意义上讲，这些顺式作用元件并不是启动子的组成部分，但却对转录有显著的影响。正如我们在第 9 章所学过的，**增强子**（enhancer）是促进基因转录的元件。相反，**沉默子**（silencer）会抑制转录。本节我们主要讨论这些元件的主要特征，在第 12 章和第 13 章介绍其作用模式。

增强子

Chambon 及同事在 SV40 早期基因的 5′-侧翼区首次发现了增强子。此前，该 DNA 区域因具有明显的 72bp 重复序列（72bp repeat）（图 10.28）而受到人们的关注。当 Benoist 和 Chambon 对这一区域进行缺失突变时，发现体内相应基因的转录效率显著降低，表明该 72bp 重复序列构成了另一个上游启动子元件。然而，Paul Berg 及同事却发现，即使方向倒转或被移到环形 SV40 基因组的另一端且与启动子相距超过 2kb，72bp 重复序列仍能激活转录。72bp 重复序列的后一种行为至少可以说明该序列并不是启动子元件。因此，把这种无方向和位置依赖性的 DNA 元件称为增强子，以区别于启动子元件。

那么，增强子如何促进转录？在第 12 章我们将了解到，增强子通过与之结合的蛋白质而发挥作用，这些蛋白质具有不同的称谓：**转录因子**（transcription factor）、**增强子结合蛋白**（enhancer-binding protein）或激活子（activator）。这些蛋白质通过与结合在启动子上的**通用转录因子**（general transcription factor）相互作用而激活转录。这种相互作用促进了前起始复合体的形成，该复合体是起始转录所必需的。因此，增强子可以使基因被激活子所诱导（或许有时被抑制）。我们将在第 11 章和第 12 章探讨这些相互作用的具体细节，同时我们还会了解到，激活子通常需要其他分子（如激素、辅激活蛋白）的协助才能发挥作用。

增强子一般位于其所调控启动子的上游，但也不绝对。事实上，早在 1983 年，Tonegawa 及同事就发现了增强子位于基因内部的一个事例。他们的研究对象是编码小鼠特殊抗体或免疫球蛋白 γ_{2b} 大亚基的基因，将该基因导入能够正常表达抗体基因但无这个特殊基因的浆细胞瘤细胞内。为检测 γ_{2b} 基因在转染细胞内的表达效率，研究者在反应体系中添加标记的氨基酸，以标记新合成的蛋白质。然后用 γ_{2b} 蛋白的抗体与标记的 $_{2b}$ 蛋白免疫共沉淀（第 5 章）。对免疫共沉淀的蛋白质进行凝胶电泳，放射自显影检测。结果表明，疑似有增强子功能的序列位于基因的一个内含子区域内，内含子是基因内的一个区域，可被正常转录，但在随后的剪接加工过程中通常会被从转录物中切除（第 14 章）。Tonegawa

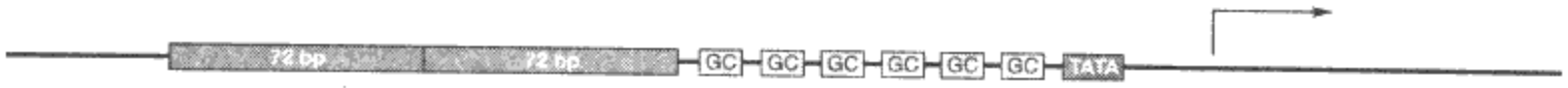


图 10.28　SV40 病毒早期基因调控区的结构。按惯例右向直角箭头表示转录的起始位点，虽然图 10.21a 所示，实际上有 3 个串联的起始位点。在转录起始位点的上游，按照从右到左的次序依次为：TATA 盒（红色）、6 个 GC 盒（黄色）、增强子（72bp 重复，蓝色）。

及同事首先对疑似增强子的区域进行缺失突变，分别删除 2 个片段，如图 10.29a 所示，然后分析 γ_{2b} 基因在转染了突变 DNA 的细胞内的表达情况。图 10.29b 显示了实验结果：尽管缺失突变发生在非编码的内含子区域，不会影响蛋白质产物的合成，但引起了蛋白质产物数量的降低，特别是进行大片段缺失（△2）时，这种情况尤为明显。

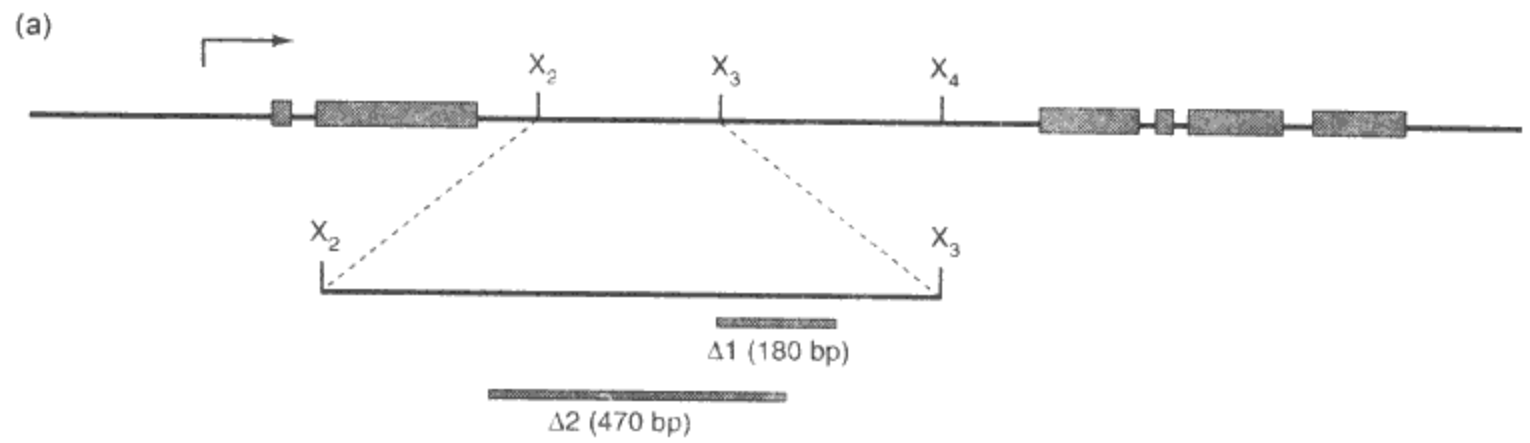


图 10.29　免疫球蛋白 γ_{2b} 重链基因增强子的缺失突变效应。（a）克隆的 γ_{2b} 基因图谱。蓝色条框代表该基因的外显子，即基因中对应于成熟 mRNA 的序列；蓝色条框间的线条代表该基因的内含子，能够被转录，但在 mRNA 前体剪接过程中被切除，不出现在成熟 mRNA 中。X_2、X_3 和 X_4 代表限制性内切核酸酶 *Xba*I 的酶切位点。Tonegawa 及同事推测增强子位于 X_2-X_3 区域。因此，研究者分别删除△1 和△2（红色条框）两个区域，分析其缺失效应。（b）在蛋白质水平上分析 γ_{2b} 基因的表达。Tonegawa 及同事分别用野生型基因（泳道 2～5）、缺失△1 区的基因（泳道 6～9）、缺失△2 区的基因（泳道 10～13）转化浆细胞；泳道 1 是未转化的对照组浆细胞。转化后，用放射性标记的氨基酸来标记新合成的蛋白质，萃取新合成的蛋白质，免疫沉淀分离 γ_{2b} 蛋白，凝胶电泳，利用荧光显影（放射自显影的一种改良方法，即在凝胶中添加荧光染料）检测放射性蛋白质。放射性辐射会引发荧光物质释放光子，可被 X 线检测到。删除△1 区后，基因表达水平略微下降，删除△2 区后，会导致基因表达水平的大幅下降。（c）γ_{2b} 基因的转录分析。Tonegawa 及同事从各转化细胞中提取 RNA，电泳后进行 Northern 印迹。泳道 1（阳性对照）是未经转化的浆细胞（MOPC 141），能够表达 γ_{2b} 基因；泳道 2（阴性对照）是未经转化的浆细胞（J558L），不能表达 γ_{2b} 基因；泳道 3～4 是经野生型 γ_{2b} 基因转化的 J558L 细胞株；泳道 5～6 是经删除△1 区的 γ_{2b} 基因转化的 J558L 细胞株；泳道 7～8 是经删除△2 区的 γ_{2b} 基因转化的 J558L 细胞株；删除△1 区会引起基因表达水平的轻微下降，删除△2 区则导致 γ_{2b} 基因表达终止。（*Source*：（*b-c*）Gillies，S. D.，S. L. Morrison，V. T. Oi，and S. Tonegawa，A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell* 33（July 1983）p. 719，f. 2&3.）

那么，这种影响是由于转录效率下降而引起的，还是另有其他原因？Tonegawa 研究小组通过 Northern 印迹实验回答了这个问题，实验所用 RNA 分别来自转染有正常或缺失突变 γ_{2b} 基因的细胞。如图 10.29c 所示，Northern 印迹结果再次证实，当疑似增强子的序列发生缺失后导致启动子的功能丧失。但这段序列是个真正的增强子吗？如果是，那么该序列在发生移动或倒转后应该仍能保持其增强转录的活性。Tonegawa 及同事首先将含有增强子的 X_2-X_3 片段倒转，如图 10.30a 中的“位置/方向 B”所示，图 10.30b 显示该增强子仍然具有功能。其次，研究者将 X_2-X_3 片段从内含子中取出，放置到启动子的上游（位置/方向 C），该增强子还能够发挥功能。在新位置处发生倒转（位置/方向 D）后还是具有功能。因此，在 X_2-X_3 片段内的某一区域具有增强子的行为：能够在启动子附近促进转录，无位置和方向依赖性。最后，研究者还比较了 γ_{2b} 基因在转化小鼠两种不同细胞内的表达情况，其中一种细胞是我们前面介绍过的浆细胞瘤细胞，另一种是成纤维细胞。结果显示，γ_{2b} 基因在浆细胞瘤细胞内的表达活性较高。这与增强子的功能相一致，因为成纤维细胞不能合成抗体，因而也就没有能激活抗体基因增强子的增强子结合蛋白。因此，抗体基因不能在成纤维细胞内表达。

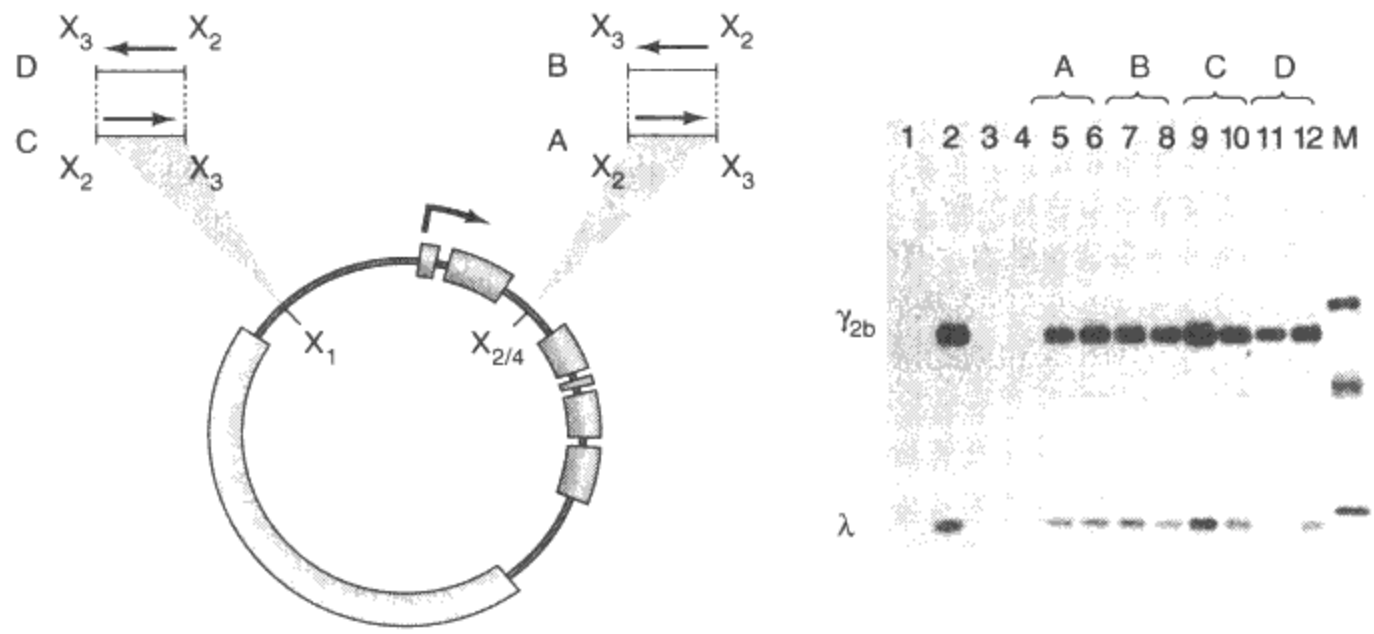


图 10.30　γ_{2b} 基因的增强子元件无方向和位置依赖性。（a）突变质粒的结构。Tonegawa 及同事将原有质粒内 γ_{2b} 基因（图 10.29）的 X_2-X_3 区切除，使增强子缺失。然后按 4 种不同的方式将 X_2-X_3 片段（包含增强子）重新插入到基因中，构建了 4 种重组质粒。质粒 A 和质粒 B：X_2-X_3 片段被正向（A）或反向（B）重新插入到原来的位置处；质粒 C 和质粒 D：X_2-X_3 片段被正向（C）或反向（D）插入到另一个 *Xba*I 酶切位点（X_1）处，此位点位于基因上游几百个碱基处。（b）实验结果。Tonegawa 及同事检测了 γ_{2b} 基因在 4 种重组质粒及其野生型质粒中的表达情况。如图 10.29b 所示，所有质粒都能高效表达 γ_{2b} 基因。泳道 1：未转染的 J558L 细胞株，缺少 γ_{2b} 基因；泳道 2～12：用不同质粒转染的 J558L 细胞株，泳道 2：未缺失 X_2-X_3 片段的野生型质粒；泳道 3～4：缺失 X_2-X_3 片段的质粒；泳道 5～6：A 质粒；泳道 7～8：B 质粒；泳道 9～10：C 质粒；泳道 11～12：D 质粒；泳道 M：蛋白质标准。（*Source*：（a）Adapted from Gillies，S. D.，S. L. Morrison，V. T. Oi，and S. Tonegawa，A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell* 33（July 1983）p. 721，f. 5.）

沉默子

增强子并不是唯一能够远程调控基因转录的 DNA 元件，沉默子也能远程调控基因的转录，只不过像其名字所寓意的那样，沉默子是抑制而不是促进基因的转录。酵母交配系统（mating system，MAT）为研究沉默子作用机制提供了一个很好的事例。酵母细胞第 III 条染色体上存在着 3 个序列十分相似的基因座：*MAT*、*HML* 和 *HMR*，但只有 *MAT* 基因座表达，*HML* 和 *HMR* 却不能表达。这是因为位于 *HML* 和 *HMR* 基因座至少 1kb 的远处存在着抑制基因活性的沉默子。我们知道，除了无活性基因的自身因素外，还有其他原因导致基因失活。因为酵母的活性基因可以被 *HML* 或 *HMR* 所

取代，而原来的活性基因被“移植”后就会失活。由此可见，这些基因受沉默子的负调控。那么，沉默子的作用机制如何？已有的数据表明，沉默子能引起染色质弯曲缠绕，产生异染色质的结构，致使转录因子无法与邻近的基因相结合，从而使基因处于失活状态。在第 13 章将对详细介绍这个过程。

有时，同一 DNA 元件既表现增强子的活性，又表现沉默子活性，这完全取决于与之结合的蛋白质的特性。例如，当甲状腺激素受体单独结合在甲状腺激素应答元件上时，此应答元件具有沉默子的功能；当甲状腺激素受体与其配体甲状腺激素结合后，再结合在甲状腺激素应答元件上时，此应答元件具有增强子的功能。在第 12 章我们将讨论该问题。

小结 增强子和沉默子都是无位置和方向依赖性的 DNA 元件，分别能够促进或抑制相关基因的转录。增强子和沉默子具有组织特异性，其活性有赖于组织特异性 DNA 结合蛋白。有时，同一 DNA 元件既表现增强子活性，又表现沉默子活性，这完全取决于与之结合的蛋白质的特性。

总结

真核生物的细胞核内存在三类 RNA 聚合酶，可通过离子交换层析的方法进行分离。RNA 聚合酶 I 存在于核仁中，另外两类 RNA 聚（RNA 聚合酶 II 和 RNA 聚合酶 III）定位在核质中。这三类细胞核 RNA 聚合酶在转录过程中行使不同的功能。RNA 聚合酶 I 负责合成较大的 rRNA（在脊椎动物中为 5.8S、18S 和 28S rRNA）前体；RNA 聚合酶 II 催化合成 mRNA 的前体 hnRNA 以及大多数的核内小 RNA（snRNA）；RNA 聚合酶 III 催化合成 5S rRNA、tRNA 的前体以及其他几种胞内小分子 RNA 和病毒 RNA。

一些真核生物的所有这 3 种细胞核 RNA 聚合酶的亚基结构均已确定。它们均由多个亚基构成，其中包括分子质量超过 100kDa 的两个大亚基。所有真核生物似乎都具有一些共有亚基，即三类 RNA 聚合酶均具有的亚基。编码酵母 RNA 聚合酶 II 全部 12 个亚基的基因均被克隆、测序，并进行了突变分析。有 3 个亚基在结构和功能上与细菌 RNA 聚合酶的核心亚基相似；有 5 个亚基是三类细胞核 RNA 聚合酶所共有的；有 2 个亚基并不是聚合酶发挥活性所必需的，至少在正常温度条件下如此；有 2 个亚基不能被归结到上述三组亚基中。

酵母 *RPB1* 基因的原初产物为 IIa 亚基。其末端功能域（CTD）是由 7 个氨基酸组成的重复序列构成，在体外，CTD 可被蛋白酶切除而使 IIa 转化为 IIb。在体内，CTD 上的 2 个丝氨酸可发生磷酸化修饰而使 IIa 转化为 IIo。含有 IIa 亚基的酶（RNA 聚合酶 IIA）负责与启动子的结合，含有 IIo 亚基的酶（RNA 聚合酶 IIO）负责转录的延伸。

原子分辨率的酵母 pol II△4/7 的结构显示，酶内具有一个很深的裂隙，能够接纳两个末端之间的线性 DNA 模板。结合有一个 Mg^{2+} 催化中心位于裂隙底部，第二个 Mg^{2+} 以较低的浓度存在，可能参与酶与核苷酸底物的结合过程。线性 DNA 模板与聚合酶复合体的几何结构显示，该复合体具有足够的空间，以利于 TFIID 与启动子 TATA 盒的结合，以及 TFIIB 将聚合酶连接到 TFIID 上，从而将聚合酶 II 定位在正确的起始位点处。

酵母 RNA 聚合酶 II（缺失 Rpb4/7 亚基）的转录延伸复合体晶体结构图像显示，钳状结构紧紧环抱着位于酶裂隙内的 RNA-DNA 杂交分子，确保转录过程中的进行性。此外，钳的 3 个环结构——舵、盖和拉链分别在下列事件中发挥重要作用：起始 RNA-DNA 杂交体的解离、维持解离状态、维持模板 DNA 的解离。酶的活性中心位于孔 1 的末端，此入口是核苷酸进入酶的通道，同时也是在聚合酶发生倒退时 RNA 逐出的通道。桥梁螺旋位于活性中心附近，该螺旋的可挠性对转录过程中的转位是至关重要的。α-鹅膏蕈碱能够抑制桥梁螺旋的弯曲性，从而阻断了转位。

Rbp1 亚基的盖可与 RNA-DNA 杂交分子发生相互作用，使杂交分子形成的碱基对不超过 8 个，盖能够与初生 RNA 链的碱基发生相互作用，使杂交分子在 8bp 区域之后处于熔解状态。Rbp1 亚基的舵通过与 DNA 模板链上的－9 位和－10 位碱基相互作用，与盖协同保持链的分离状态。Rbp2 亚基的叉状环 1 可与 RNA 链的－5 位、－6 位和－7 位碱基发生相互作用，使 RNA-DNA 杂交分子的长度维持在 8 个碱基对左右。

在通过入口移向 RNA 聚合酶 II 活性位点的过程中，正被掺入的核苷酸必须先进入 E（进入）位点，而 A 位点是磷酸二酯键形成的位点。相对于在 A 位点的位置，该核苷酸在 E 位点发生了倒转。E 和 A 位点具有一定的重叠，在 E 和 A 位点之间核苷酸发生的旋转可能在碱基和糖的特异性识别方面发挥作用。活性位点结合有 2 个金属离子（Mg^{2+} 或 Mn^{2+}），其中一个会永久地结合在酶的活性位点上，而另一个进入活性位点，与正被掺入的核苷酸结合形成复合物。

包含 12 个亚基的 RNA 聚合酶 II 的结构显示，Rbp4/7 的存在会迫使钳处于闭合状态，由于转录起

始由 12 亚基的聚合酶引发，其钳状结构处于闭合状态，所以，模板 DNA 链在进入活性位点之前，启动子 DNA 要先行熔解。该结构也显示，Rbp4/7 延伸到聚合酶的平台区，有助于特定通用转录因子的结合，从而促进转录的起始。

II 类启动子可能包含有 5 个保守元件：由 4 个元件构成的核心启动子及上游启动子元件。4 个核心启动子元件包括 TFIIB 识别元件（BRE）、TATA 盒、起始子（Inr）和下游启动子元件（DPE）。在大多数启动子中，至少会缺失其中的 1 个元件。事实上，缺失 TATA 盒的启动子通常都含有 DPE，至少在果蝇中情况是这样的。起始特异性基因高效表达的启动子一般都含有 TATA 盒，而起始持家基因表达的启动子一般缺失 TATA 盒。上游启动子元件通常存在于 II 类核心启动子的上游，与核心启动子不同的是，上游启动子元件能够与基因特异性转录因子结合。例如，GC 盒与转录因子 Sp1 结合，而 CCAAT 盒与转录因子 CTF 结合。与核心启动子不同，上游启动子元件无方向依赖性，但与典型的增强子相比，上游启动子元件具有位置依赖性。

在不同的物种间，I 类启动子序列并不具有很好的保守性，但在一般结构上还是具有较高的保守性。I 类启动子由两个元件构成：包围着转录起始位点的核心元件和大约在上游 100bp 远处的上游启动子元件（UPE）。这两个元件间的空间距离是十分重要的。

RNA 聚合酶 III 转录一套较短的基因，典型 III 类基因（I 型和 II 型）启动子的全部序列均位于基因内部。III 类基因（5S rRNA 基因）的 I 型内部启动子可分为 3 个区域：A 盒、短的中间元件、C 盒。III 类基因（如 tRNA 基因）的 II 型内部启动子可分为两个区域：A 盒与 B 盒。其他 III 类基因（如 7SK 基因和 U6 RNA 基因）的启动子称为 III 型启动子，缺乏内部启动子，但在基因的 5′-侧翼区具有与 II 类启动子十分相似的启动子区域，且包含有 TATA 盒。

增强子和沉默子都是无位置和方向依赖性的 DNA 元件，分别能够促进或抑制相关基因的转录。增强子和沉默子具有组织特异性，其活性有赖于组织特异性 DNA 结合蛋白。

复习题

1. 图示说明真核生物细胞核 RNA 聚合酶在进行 DEAE- Sephadex 层析分析时，从层析柱中洗脱的情况。如果在 1μg/mL α-鹅膏蕈碱存在的情况下，用相同的组分进行实验，会出现怎样的结果。

2. 描述能够证明 RNA 聚合酶 I 主要定位在细胞核的核仁内的实验方案，并给出实验结果。

3. 描述能够证明 DNA 聚合酶 III 负责转录合成 tRNA 和 5S rRNA 的实验方案，并给出实验结果。

4. RNA 聚合酶 II 由多少个亚基组成？哪些为核心亚基？哪几个是所有 3 种细胞核 RNA 聚合酶共有的亚基？

5. 描述利用抗原决定簇标签法从酵母中纯化 RNA 聚合酶 II 的实验过程。

6. 在制备的 RNA 聚合酶 II 样品中，其最大亚基（RPB1）有 3 种不同的形式。请给出这 3 种不同形式亚基的名称，并指出它们在 SDS-PAGE 上的相对位置。这些不同形式的 RPB1 之间具有怎样的区别？请列举相关证据。

7. RPB1 亚基的 CTD 有什么结构特点？

8. 简图说明酵母 RNA 聚合酶 II 的结构，并显示 DNA 的结合位点且列举相关证据，同时也要显示活性位点的位置。

9. 在 RNA 聚合酶的催化中心有多少 Mg^{2+} 参与了催化反应？为什么在酵母 RNA 聚合酶 II 的晶体结构中很难被检测到其中的一个 Mg^{2+} 的存在？

10. 引用相关的证据，说明在 RNA 聚合酶 II 倒退时，孔 1 也是逐出 RNA 的通道。

11. 进行性转录的含义是什么？RNA 聚合酶 II 的哪一部分能够确保转录的进行性？

12. RNA 聚合酶 II 的盖、舵及叉环 1 可能的功能是什么？根据最新的研究数据做出回答。

13. 桥梁螺旋可能的功能是什么？α-鹅膏蕈碱与该功能间的关系如何？

14. RNA 聚合酶 II 的 A 位点和 E 位点的含义是什么？在核苷酸选择过程中的作用是什么？

15. Rpb4/7 复合体在 RNA 聚合酶 II 的钳状结构的打开或闭合过程中具有怎样的功能？其证据是什么？

16. 包含 12 亚基的 RNA 聚合酶 II 可与启动子 DNA 发生相互作用，这暗示启动子 DNA 必须处于什么状态才能与聚合酶发生相互作用？

17. 图例说明 RNA 聚合酶 II 的启动子结构特点，并显示所有类型的元件。

18. 哪类基因可能具有 TATA 盒？而哪类基因又可能缺失 TATA 盒？

19. TATA 盒与 UPE 之间可能存在怎样的关联？

20. 如果将 TATA 盒从 II 类启动子中切除后，最有可能产生的两个效应是什么？

21. 图例说明接头扫描的实验过程，通过接头扫描实验能够给出怎样的信息？

22. 列举 II 类启动子的两个常见上游元件，它

们与核心启动子元件相比，具有哪些不同点？

23. 图例说明典型 I 类启动子的结构特点。

24. I 类启动子的元件是如何被发现的？请列举实验结果。

25. 设计实验方案，证明维持 I 类启动子的两个元件之间适宜的空间距离的重要性，并给出实验结果。

26. 图示比较典型与非典型 III 类基因启动子的差异，并举例说明。

27. 描述定位 5S rRNA 基因启动子 5′-边界的实验方案，并给出实验结果。

28. 解释增强子具有组织特异性的事实。

分析题

1. II 类基因转录的起始位点鸟嘌呤位于 TATA 盒最后一个碱基下游 30bp 处。如果在两者之间删除长度为 20bp 的片段，然后用突变的 DNA 转化细胞，转录的起始位点还是同一个鸟嘌呤吗？如果不是，那么转录的起始位点在哪里？你如何确定转录的起始位点？

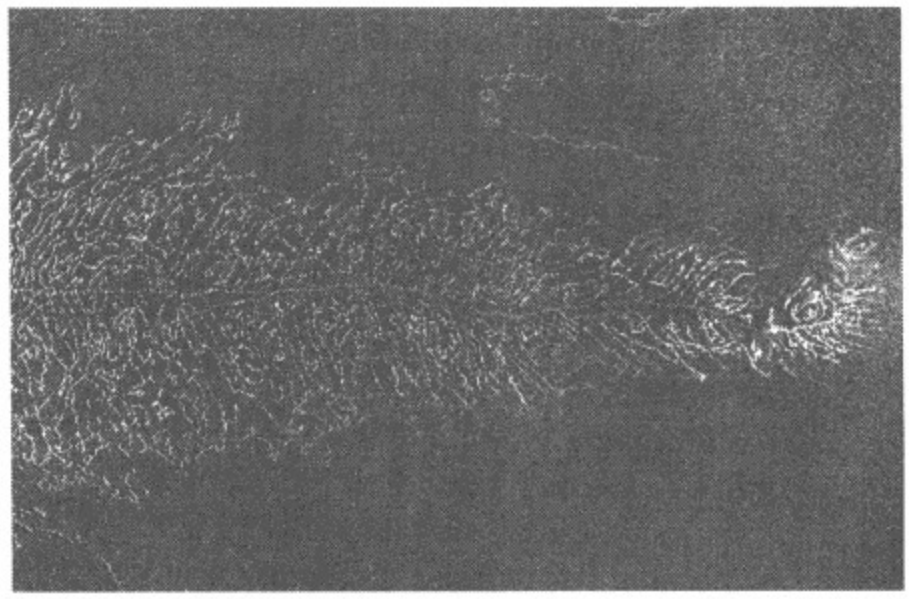
2. 你推测位于基因上游的一个重复序列可能具有增强子的功能，描述能够证明你的推断的实验方案，并预测实验的结果。你的实验应该确实证明该序列具有增强子的功能，而不具有启动子的功能。

3. 你正在研究一个新的 II 类启动子，但没有发现已熟知的序列，描述确定启动子序列位置的实验方案，并简要阐述实验结果。

4. 阐述用于确定 5S rRNA 基因启动子 3′端的引物延伸分析实验过程。

翻译　徐启江　校译　李玉花

第 11 章　真核生物中的通用转录因子



由 DNA 模板合成 RNA 的假彩色透射电镜照片，呈羽毛状结构。羽毛的主脉是双链 DNA 模板，其上包被着蛋白质，大量 RNA 分子从 DNA 模板上成簇延伸出来（×14 000）。（*Source*：© *Oscar L. Miller/SPL/Photo Researchers*，*Inc.*）

与细菌 RNA 聚合酶相比，真核生物的 RNA 聚合酶不能独立结合到启动子上，需要依赖称作转录因子的蛋白质引导才能发挥作用。转录因子分为两种，**通用转录因子**（general transcription factor）和**基因特异性转录因子**（gene-specific transcription factor）（**激活因子**，activator）。通用转录因子可将 RNA 聚合酶引导至相应的启动子上，但程度较弱，仅支持本底水平的转录。此外，单靠通用转录因子和 3 种聚合酶只能实施最小量转录调控，而实际上细胞对转录的调控极其精细。不过，通用转录因子使 RNA 聚合酶结合至相应的启动子上的作用，不仅是绝对重要的，而且由于很多多肽的参与也是非常复杂的。本章将学习与 3 种 RNA 聚合酶及其启动子相互作用的通用转录因子。

11.1　II 类转录因子

通用转录因子与 RNA 聚合酶结合形成**前起始复合物**，只要有核苷酸，该复合物结合就能起始转录。两者的紧密结合伴随着开放启动子复合物的形成，使 DNA 从转录起始位点开始解链，以便聚合酶读取。我们首先从含有聚合酶 II 的前起始复合物的装配开始，虽然过程很复杂，但却是研究最清楚的部分。一旦了解了 II 类通用转录因子的作用机制，I 类和 III 类的机制就容易理解了。

II 类前起始复合物

II 类前起始复合物包括聚合酶 II 和 6 种通用转录因子，分别为 **TFIIA、TFIIB、TFIID、TFIIE、TFIIF 和 TFIIH**。许多研究表明至少在体外，II 类通用转录因子与 RNA 聚合酶 II 以一定顺序结合到形成中的前起始复合物上。Danny Reinberg、Phillip Sharp 及同事通过 DNA 凝胶阻滞、DNase 和羟基自由基足迹实验（第 5 章），确定了 II 类前起始复合物中多数因子的结合顺序。

图 11.1a 显示 Danny Reinberg 和 Jack Greenblatt 及同事用 TFIIA、TFIID、TFIIB、TFIIF 及 RNA 聚合酶 II 进行了凝胶阻滞实验的结果。有 4 种明显不同的复合物存在（图 11.1 左侧）。在含有腺病毒主要晚期基因启动子的 DNA 中添加 TFIID 和 TFIIA 时，形成 DA 复合物（泳道 1）。同时添加 TFIIB、TFIID 和 TFIIA 时，形成 DAB 复合物（泳道 2）。图的中间部分显示了在 DAB 复合物中添加不同浓度的 RNA 聚合酶 II 和 TFIIF 后的情况。泳道 3 加入了标记的 TFIID＋A＋B＋F 全部 4 种因子，但没有 RNA 聚合酶。这 4 种因子形成的复合物与 DAB 复合物之间未检测到差异。因此，TFIIF 似乎不能独立结合到 DAB 上。但是逐渐增加聚合酶的量时（泳道4～7），两种新复合物出现了。由此可以推测两种复合物中含有聚合酶和 TFIIF，所以最上面的复合物称为 DABPolF 复合物，另一个新复合物（DBPolF）

由于缺少 TFIIA 而迁移较快。当添加了过量聚合酶从而形成最大量 DABPolF 复合物后，开始降低 TFIIF 的量（泳道 8～11），直至降为零。结果 DABPolF 复合物的形成量随之减少直至消失，所以即使含有大量聚合酶但没有 TFIIF 时（泳道 12），DABPolF（或 DABPol）复合物也不能形成。以上结果表明 RNA 聚合酶和 TFIIF 因子需同时加入才能形成前起始复合物。

Reinberg、Greenblatt 及同事利用同样的凝胶阻滞实验测定不同因子的结合顺序，每次都去掉一个或多个因子。最极端的例子在泳道 13 中，标记为－D。结果显示如果去掉 TFIID，即使所有其他因子都存在也不会形成任何复合物。这种对 TFIID 的依赖性支持了 TFIID 是第一个结合因子的假说，其他因子的结合依赖于 TFIID 首先与 TATA 盒结合。泳道 14 表明去掉 TFIIB 时，仅有 DA 复合物形成。因此聚合酶和 TFIIF 的结合需要 TFIIB。泳道 15 表明去掉 TFIIA 没有什么变化，因此至少在体外 TFIIA 不重要，而且该泳道中条带的位置与两个大复合物中的较小者有相同迁移率，提示这个较小的复合物为 DBPolF。最后一个泳道含有所有蛋白质因子，呈现两个大复合物及剩余的 DAB 复合物。

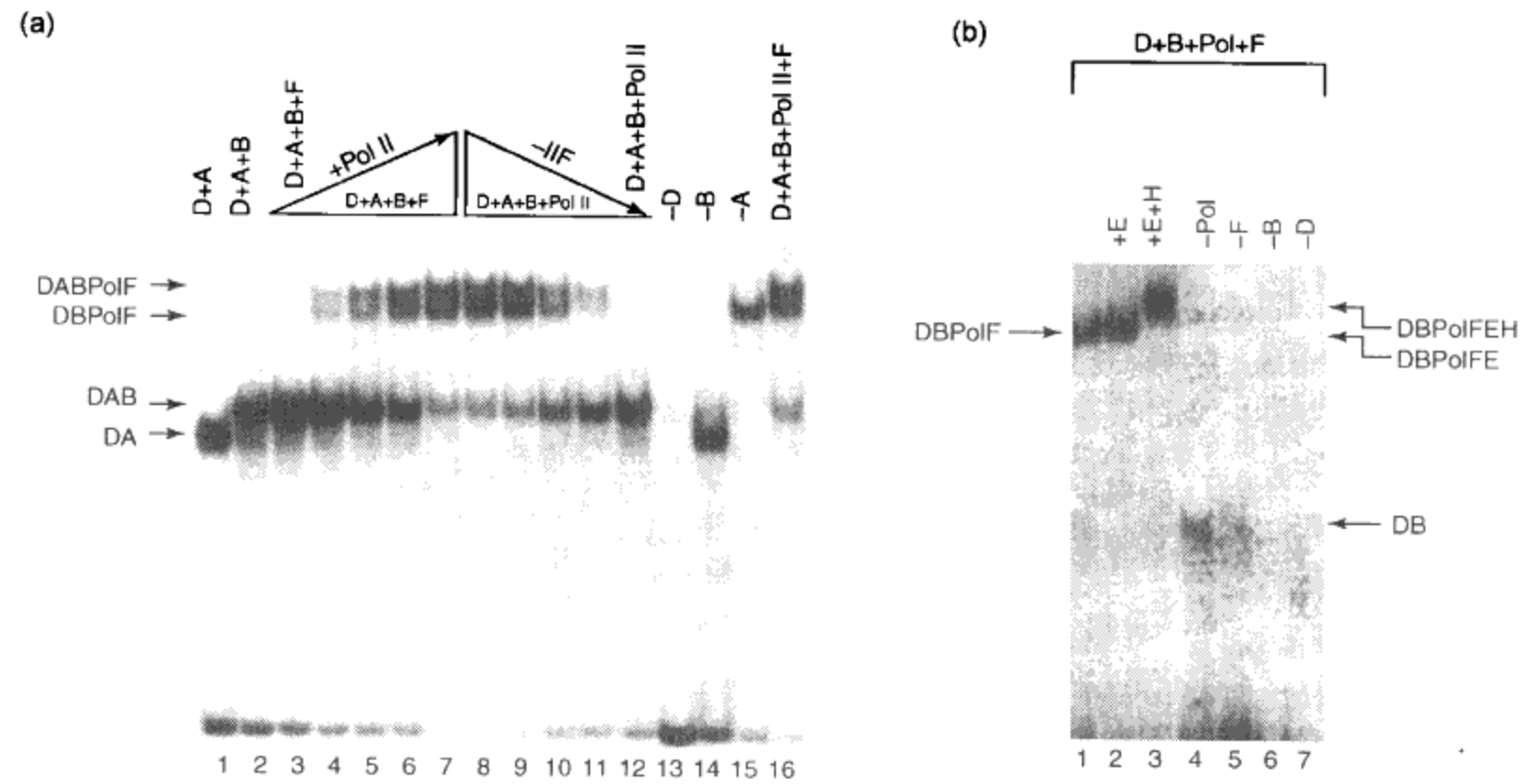


图 11.1　前起始复合物的构建。（a）DABPolF 复合物。Reinberg 及同事用 4 种 II 类因子（TFIID、TFIIA、TFIIB、TFIIF）及 RNA 聚合酶 II，分别与含有标记腺病毒主要晚期基因启动子的 DNA 进行凝胶阻滞实验。泳道 1 显示 DA 复合物，由 TFIID 和 TFIIA 形成。泳道 2 证明加入 TFIIB，形成新复合物 DAB。泳道 3 含有 TFIID、TFIIA、TFIIB 和 TFIIF，但与泳道 2 的结果完全相同。因此，在无聚合酶 II 时 TFIIF 不结合 DAB 复合物。泳道 4～7 显示 4 种转录因子再加越来越多的聚合酶 II，出现越来越多的大复合物（DABPolF 和 DBPolF）。泳道 8～11 含有越来越少的 TFIIF，可见大复合物越来越少。泳道 12 显示无 TFIIF 时，基本上不形成 DABPolF 和 DBPolF。因此，TFIIF 可能将聚合酶 II 引入复合物。右边的泳道显示每次去掉一个因子时的结果。泳道 13 无 TFIID，不能形成复合物。泳道 14 显示无 TFIIB 时，只形成 DA 复合物。泳道 15 证明无 TFIIA 时，仍能形成 DBPolF。泳道 16 显示当所有因子存在时，所有大复合物都出现了。(b) DBPolFEH 复合物。将 DBPolF 复合物（缺 TFIIA，泳道 1）装配到含有标记腺病毒主要晚期基因启动子的 DNA 上，接着依次添加 TFIIE、TFIIH，进行凝胶阴滞实验。随着新转录因子的添加，复合物的分子质量逐渐增大，迁移率逐渐降低。在右边标出了两种复合物的迁移率。泳道4～7再次显示去掉各因子的结果，具体因子于每一泳道的顶部标出。这里最好的结果是仅有 DB 复合物形成，最差是无 TFIID 时没有任何复合物形成。（*Sources*：(a)Flores，O.，H. Lu，M. Killeen，J. Greenblatt，Z. F. Burton，and D. Reinberg，The small subunit of transcription factor IIF recruits RNA polymerase II into the preinitiation complex. *Proceedings of the National Academy of Sciences USA*，88（Nov 1991）p. 10001，f. 2a.（b）Cortes，P.，O. Flores，and D. Reinberg. 1992. Factors involved in specific transcription by mammalian RNA polymerase II：Purification and analysis of transcription factor IIA and identification of transcription factor IIJ. *Molecular and Cellular Biology* 12：413-21. American Society for Microbiology.）

1992 年，Reinberg 及同事继续研究了另外两个转录因子 TFIIE 和 TFIIH。图 11.1b 证明可以从 DBPolF 复合物开始，然后依次加入 TFIIE 和 TFIIH，随之产生较大的复合物，并且随着每个因子的加入迁移率降低，最终在此实验中形成了前起始复合物 DBPolFEH。后 4 个泳道再次表明去掉任何一个早期因子（聚合酶 II、TFIIF、TFIIB 或 TFIID）都会阻断完整前起始复合物的形成。

因此，各转录因子（和 RNA 聚合酶）在体外添加到前起始复合物的顺序如下：TFIID（或 TFIIA＋TFIID）、TFIIB、TFIIF＋RNA 聚合酶 II、TFIIE、TFIIH。确定了体外大多数通用转录因子（及 RNA 聚合酶）结合到前起始复合物的顺序后，将研究重点转移到每个因子在 DNA 链上的具体结合位置。从 Sharp 小组开始，几个研究小组利用足迹法对此进行了研究。图 11.2 为 DA 和 DAB 复合物足迹实验结果。Reinberg 与同事用两种不同试剂切割蛋白质-DNA 复合物，一种是能产生羟基自由基的 1,10-二氮杂菲（OP)-铜离子复合物（两图中的1～4 泳道）；另一种是 DNase I（两图中的 5～8 泳道）。图 11.2a 显示模板链的结果，图 11.2b 显示非模板链的结果。图 11.2a 中泳道 3 和 7 表明 TFIID 和 TFIIA 保护了 TATA 盒区。图 11.2b 中泳道 3 和 7 表明 DA 复合物也保护了非模板链的 TATA 盒区。图 11.2a 中泳道 4 和 8 显示，添加 TFIIB 形成 DAB 复合物对模板链的足迹没有改变。在非模板链上的结果基本相同，但有细微差异。如泳道 8 所示加入 TFIIB 使＋10 位 DNA 对 DNase 更敏感。因此，TFIIB 看起来没有较大区域地覆盖 DNA，而是干扰了 DNA 结构使其对 DNase 攻击的敏感程度发生了改变。

RNA 聚合酶 II 是一个很大的蛋白质，因此预测它能覆盖很长一段 DNA 区域，留下一个很大的足迹，图 11.3 证实了这一预测。TFIID、TFIIA 及 TFIIB 保护 DAB 复合物中的 TATA 盒区（－17～－42bp)，RNA 聚合酶 II 和 TFIIF 在非模板链上又增加了 34bp 的保护区（从－17bp 到大约＋17bp)。图 11.4 显示 TFIIF 在形成 DABPolF 复合物中的作用。聚合酶 II（红色）和 TFIIF（绿色）协同作用，可能通过形成二元复合物结合到已形成的 DAB 复合物中。

小结　在体外，各转录因子以下列顺序结合到 II 类启动子上（包括腺病毒主要晚期基因启动子)：①TFIID 在 TFIIA 的协助下，结合到 TATA 盒，形成 DA 复合物；②接着是 TFIIB 的结合；③TFIIF 协助 RNA 聚合酶结合到－34～＋17bp的 DNA 区域。剩余的因子按 TFIIE、TFIIH 顺序结合，最后形成 DABPolFEH 前起始复合物。在体外，TFIIA 对复合物的形成是可选的。

TFIID 的结构和功能

TFIID 是一个含有 **TATA 盒结合蛋白（TATA-box-binding protein，TBP）** 和 8～10 个 **TBP 相关因子**（TBP-associated factor，**TAF**，具体为 $\mathbf{TAF_{II}}$）的复合物。这里必须标注“II”，因为 TBP 也参与 I 类和 III 类基因的转录，并且分别与 I 类和 III 类前起始复合物中的不同 TAF（TAF_I 和 TAF_{III}）结合。我们将在本章的后面讨论在 I 类和 III 类启动子转录中 TBP 和 TAF 的作用。先讨论 TFIID 的组分及其相应活性，从 TBP 和 TAF_{II} 开始。

TATA 盒结合蛋白　TBP 是第一个在 TFIID 复合物中鉴定出来的多肽，在进化上高度保守，相距很远的不同物种，如酵母、果蝇、植物和人类，TATA 盒结合域的氨基酸序列有 80％以上的相似性，这些区域包含该物种 TBP 蛋白羧基端的 180 个富含碱性的氨基酸。另一个进化保守性表现为酵母 TBP 在哺乳动物通用转录因子形成的前起始复合物中功能正常。

Tjian 小组通过 DNase I 足迹实验证明羧基端的 180 个氨基酸对 TBP 的功能非常重要。将重组的仅含 180 个羧基端氨基酸的人类 TBP 与启动子的 TATA 盒结合，仍能像天然 TFIID 因子一样有效。

TFIID 中的 TBP 如何结合到 TATA 盒呢？起初认为是像其他大多数 DNA 结合蛋白（第 9 章）一样，与 TATA 盒大沟中的碱基发生特异性作用，后来证明这一假说是错误的。以 Diane Hawley 和 Robert Roeder 为首的两个小组证实 TFIID 中的 TBP 结合在 TATA 盒的小沟。

Barry starr 和 Hawley 改变了 TATA 盒的所有碱基，使大沟改变，但小沟不变。这种可能性在于小沟中次黄苷（I）的次黄嘌呤与腺嘌呤（A）类似，胞嘧啶与胸腺嘧啶类似，但是在大沟中却完全不同（图 11.5a)。因此，将腺病毒主要晚期基因 TATA 盒的胸腺嘧啶换成胞嘧啶，腺嘌呤换成次黄苷（TATAAAA 换成 CICIII，图 11.5b)，然后通过 DNA 迁移率变动实验检测结合到 CICI 框和标准 TATA 盒上的 TFIID 的迁移率。如图 11.5c 所示：CICI 框与 TATA 盒具有完全相同的功能，而无关的非特异性 DNA 根本不能与 TFIID 结合。因此，只要小沟不改变，TATA 盒的碱基改变不影响 TFIID 的结合，这是 TFIID 与 TATA 盒小沟结合而不是大沟的强有力的证据。

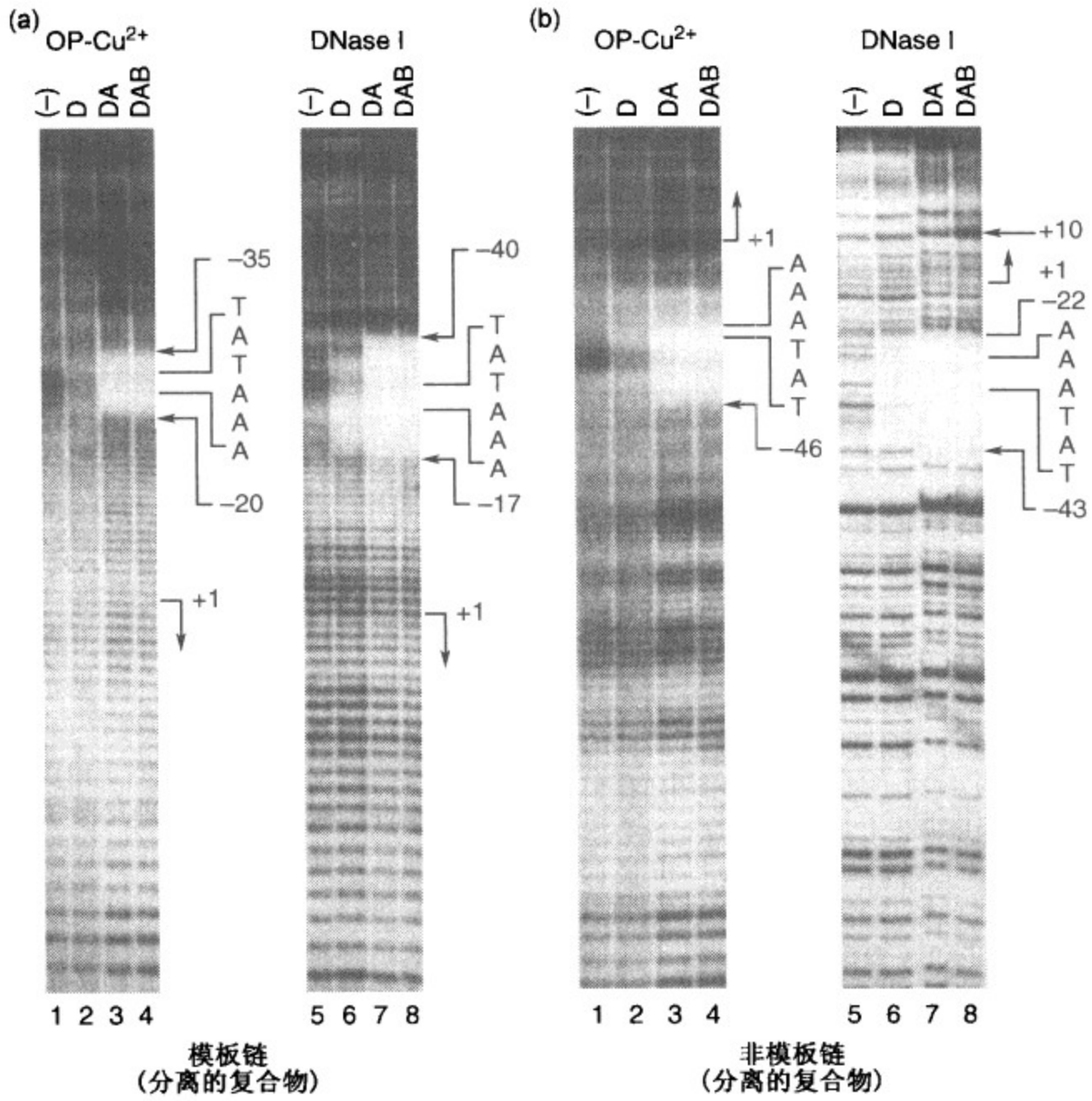


图 11.2　DA 和 DAB 复合物的足迹实验。Reinberg 及同事利用 DNase 和 DNA 链裂解剂（DNA strand breaker）—1,10-二氮杂菲-铜离子复合物（OP-Cu^{2+}）进行 DA 和 DAB 复合物的足迹实验。（a）模板链上的足迹。在 TATA 盒区，DA 和 DAB 复合物形成（TATAAA 在右边由上而下标出）。（b）非模板链上的足迹。同样，DA 和 DAB 复合物所保护区域仍以 TATA 盒为中心（TATAAA 在右边由上到下标出）。右上部箭头表示在＋10 碱基处有一增强的 DNA 切割位点。（*Source*：Adapted from Maldonado E.，I. Ha，P. Cortes，L. Weiss，and D. Reinberg，Factors involved in specific transcription by mammalian RNA polymerase II：Role of transcription Factors IIA，IID，and IIB during formation of a transcription-competent complex. *Molecular and Cellular Biology* 10（Dec 1990）p. 6344，f. 9. American Society for Microbiology.）

那么 TFIID 如何与 TATA 盒的小沟结合呢？Nam-Hai Chua、Roeder、Stephen Burley 及同事利用拟南芥（*Arabidopsis thalliana*）TBP 的晶体结构解决了这一问题。他们发现 TBP 的晶体结构像马鞍，有两个“马镫”，让人自然地联想到 TBP 像马鞍固定在马背上那样结合在 DNA 上，其两侧的镫子使 TBP 的结构大致保持对称的“U”形。而后在 1993 年，Paul Sigler 小组和 Stephen Burley 小组独立解析出与人工合成小片段双链 DNA 结合的 TBP 的晶体结构，该 DNA 片段包含有 TATA 盒。他们发现 TBP 并不是像马鞍那样被动地结合在 DNA 上。图 11.6 显示这一结构。马鞍弯曲的下部与 DNA 并不十分吻合，而是沿 DNA 的长轴排开，其曲度迫使 DNA 弯曲 80°，这一弯曲伴随着 DNA 螺旋的变形使小沟张开。这一结构在 TATA 盒的第一个和最后一个碱基处最明显（碱基对 1 和 2 之间、7 和 8 之间）。在这两个位点处 TBP 马镫的两个苯丙氨酸侧链插入两个碱基对之间，引发 DNA 扭结。这一变形有助于解释为什么此处的 TATA 序列如此保守：与其他二核苷酸键相比，DNA 双螺旋中的 T-A 键相对容易断裂，由此推测 TATA 盒的变形对转录起始很重要。事实上也很容易想到 DNA 小沟

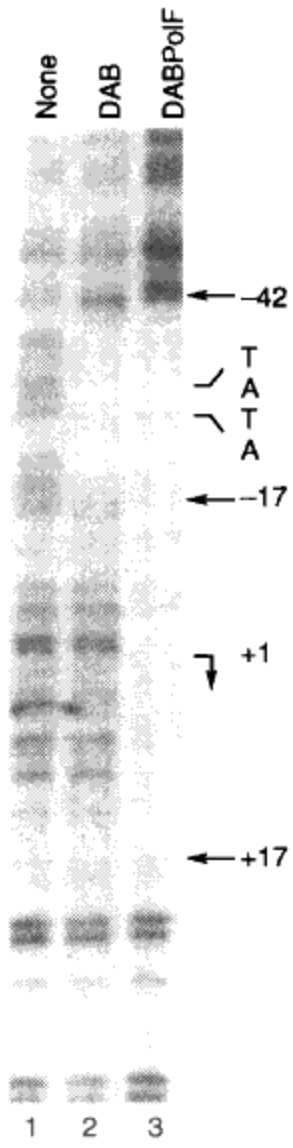


图 11.3　DABPolF 复合物足迹。Reinberg 及同事用 TFIID、TFIIA 和 TFIIB（泳道 2），TFIID、TFIIA、TFIIB 和 TFIIF 及 RNA 聚合酶 II（泳道 3）进行 DNA 酶足迹。当复合物中加入 RNA 聚合酶和 TFIIF 时，足迹的范围扩大至＋17bp 位置。这与 RNA 聚合酶 II 的大小一致。（*Source*: Flores O., H. Lu, M. Killeen, J. Greenblatt, Z. F. Burton, and D. Reinberg, The small subunit of transcription factor IIF recruits RNA polymerase II into the preinitiation complex. *Proceedings of the National Academy of Sciences USA* 88 (Nov 1991) p. 10001, f. 2b.）

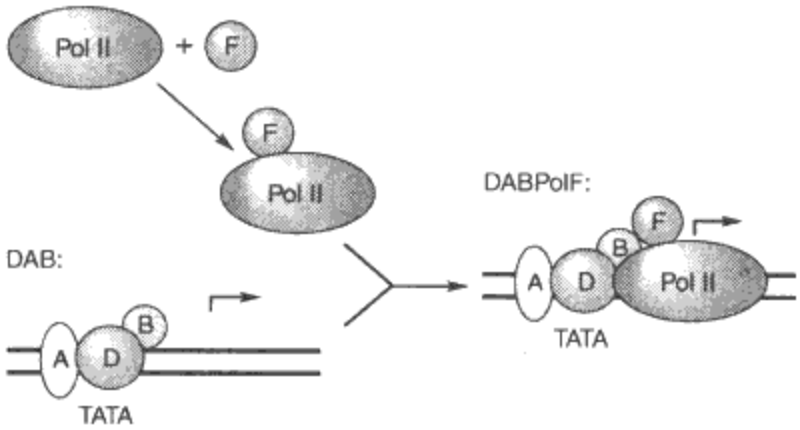


图 11.4　DABPolF 复合物形成的模型。聚合酶 II（Pol II，红色）通过与 TFIIF（绿色）结合，再结合到 DAB 复合物上，形成 DABPolF 复合物，表明聚合酶在 DNA 下游延伸 DAB 足迹。所以，它结合在 TFIID、TFIIA 和 TFIIB 结合位点（TATA 盒）下游的 DNA 上。

的张开有助于 DNA 双链的解离，这是形成开放启动子复合体的重要部分。

小结　TFIID 含有 38kDa 的 TATA 盒结合蛋白（TBP），以及若干多肽，即 TBP 相关因子（TAF_{II}）。人类 TBPC 末端的 180 个氨基酸是 TATA 盒结合域。TBP 和 TATA 盒在 DNA 小沟发生相互作用。马鞍形 TBP 与 DNA 呈直线排列，鞍的下部迫使小沟张开，使 TATA 盒呈 80°弯曲。

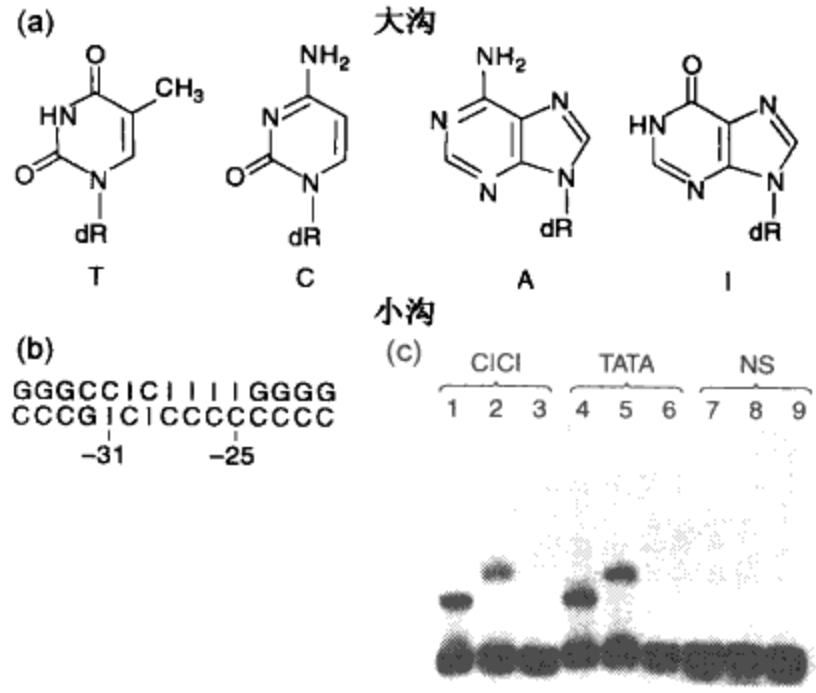


图 11.5　C 代替 T、I 代替 A 对 TFIID 结合 TATA 盒的影响。（a）大沟与小沟中的核苷酸结构。注意小沟中的 T 与 C 类似（绿色，下方），而大沟中的差别很大（红色，上方）。同上，小沟中 A 和 I 类似，而在大沟中差异很大。（b）腺病毒主要晚期启动子（MLP）的 TATA 盒中 T 被 C 取代，A 被 I 取代，形成 CICI 框。（c）TBP 结合 CICI 框。用不同 DNA 片段进行凝胶阻滞实验，DNA 片段分别为包含有 CICI 框的 MLP（泳道 1～3），或正常 TATA 盒（泳道 4～6），或不带启动子元件的无关非特异性 DNA（NS）（泳道 7～9）。每组的第一泳道（1、4、7）为酵母 TBP。每组的第二泳道（2、5、8）为人 TBP。每组的第三泳道为缓冲液。酵母和人 TBP 所产生的蛋白质-DNA 复合体大小稍有差异，但是以 CICI 框取代 TATA 盒对形成蛋白质-DNA 复合体几乎没有影响。因此，TBP 对 TATA 盒的结合并未因 CICI 框的替换而减弱。（*Source*:（*b-c*）Starr, D. B. and D. K. Hawley, TFIID binds in the minor groove of the TATA box. *Cell* 67 (20 Dec 1991) p. 1234, f. 2b. Reprinted by permission of Elsevier Science.）

TBP 的多功能性　分子生物学充满了神奇，其中之一就是 TATA 结合蛋白（TBP）的多功能性。TBP 不仅与聚合酶 II 的含 TATA 盒启动子作用，而且也与无 TATA 盒的启动子作用。更为惊人的是它还与无 TATA 盒的聚合酶 I 和 III 所识别的启动子作用。换句话说，TBP 对所有聚合酶和启动子类型都起作用，是普适的真核转录因子。

Ronald Reeder、Steven Hahn 及同事在利用温度敏感 TBP 对突变酵母的研究中验证了 TBP 的普适性。预测升温将阻断突变体中聚合酶 II 的转录，但结果显示突变体的聚合酶 I 和 III 的转录也受到严

图 11.6　TBP-TATA 盒复合物的结构。以 Sigler 及同事解析的 TBP-TATA 盒复合物的晶体结构为基础，顶部显示 TBP 的骨架（橄榄色）。马鞍的长轴在纸平面上，位于 TBP 下方的 DNA 以多种颜色显示。与蛋白质作用的 DNA 骨架为橙色，碱基对为红色。注意蛋白质如何打开 DNA 小沟并将该区域的螺旋拉直的。在图中央的右侧，TBP 的一个环状马镫（橄榄色）插入小沟中。另一个马镫有相同功能，但在 DNA 后部无法看到。不与 TBP 作用的 DNA 两端以蓝色和灰色表示，蓝色为骨架，灰色为碱基对。DNA 的左端伸出纸平面 25°，右端插入纸平面 25°。由 TBP 引起的整个 DNA 约 80°弯折显而易见。（*Source*：Klug, A. Opening the gateway. *Nature* 365（7 Oct 1993）p. 487，f. 2. © Macmillan Magazines Ltd.）

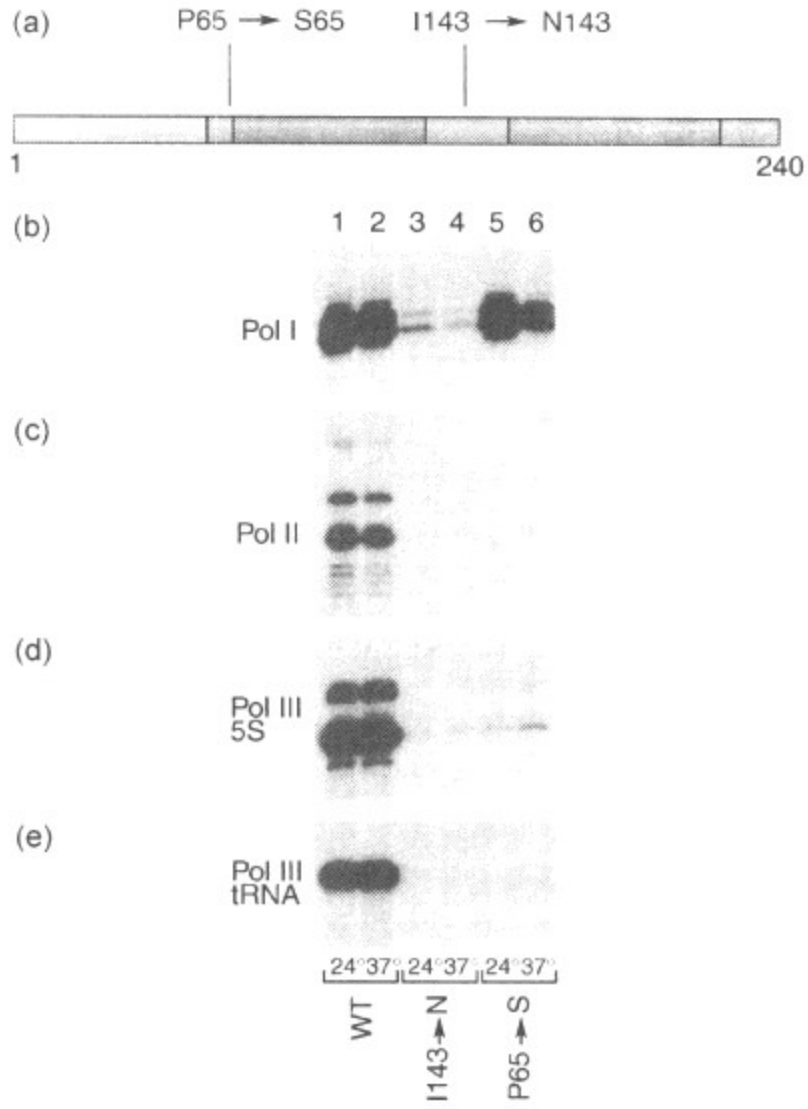


图 11.7　TBP 的突变对 3 种 RNA 聚合酶转录的影响。（a）突变的位置。蓝色和红色区域表明 TBP 保守的 CTD。红色区域表示与 DNA 结合的两个重复元件。一个突变是 P65→S65，即脯氨酸 65 突变为丝氨酸。另一个是 I143→N143，即异亮氨酸 143 突变为天冬氨酸。（b～e）突变的效应。野生型和突变酵母的提取物在 37℃进行热激或放置在 24℃，如图底部所示。S1 分析检测提取物在 3 种核 RNA 聚合酶识别的启动子上起始转录的能力。（b）rRNA 启动子（聚合酶 I）。（c）CYC1 启动子（聚合酶 II）。（d）5S rRNA 启动子（聚合酶 III）。（e）tRNA 启动子（也是聚合酶 III）。即使无热激，I143→N143 提取物从 4 个启动子上开始转录的效率也很低。P65→S65 提取物在聚合酶 II 和 III 启动子上转录有缺陷，但即使在热激之后仍识别聚合酶 I 所识别的启动子。（*Source*：（*a*）Adapted from Schultz, M. C., R. H. Reeder, and S. Hahn. 1992. Variants of the TATA binding protein can distinguish subsets of RNA polymerase I, II, and III promoters. *Cell* 69：697-702.）

重影响。

图 11.7 显示了这一假设的证据从野生型和两种对温度敏感的（TBP 有损伤的）突变体中制备了无细胞提取物如图 11.7a 所示。对细胞进行两组处理：一组是 24℃生长，37℃热激 1h；另一组是在低温下生长。然后加入 DNA 片段，其中含有被所有 3 种聚合酶识别的启动子，S1 核酸酶作图测定转录。图 11.7b～e 为实验结果。正如预期的，泳道 1 和 2 显示热激对野生型提取物没有影响。无论是否热激，I143→N143 突变体的提取物几乎不能使任何一种聚合酶转录（泳道 3 和 4）。很明显，TBP 突变不仅影响聚合酶 II 的转录，也影响其他两种聚合酶的转录。突变体 P65→S65 显示在聚合酶 I 与其他两种聚合酶活性之间的有趣差异。其提取物无论是否经热激处理，几乎都不支持 II 类和 III 类基因的转录。而无热激处理的提取物可使聚合酶 I 的转录达到野生型水平，热激处理的提取物使聚合酶 I 的转录效率降低约 2 倍。最终，加入野生型 TBP 可恢复 3 种聚合酶在突变提取物中的转录活性（结果未给出）。

TBP 不仅普遍参与真核细胞的转录，还可能参与一类完全不同的生物古菌（archaea）的转录。古菌（以前称古菌，archaebacteria）是一类没有细胞核的单细胞生物，通常生活于极端环境中，如温泉或深海沸泉喷发口。它们与细菌和真核生物都不同，而在许多方面更像真核生物。1994 年，Stephen

Jackson 及同事报道了一种古菌，伍斯氏火球菌（*Pyrococcus woesei*），它能产生一种在结构和功能上类似于真核 TBP 的蛋白质。推测这种蛋白质参与识别 TATA 盒，定位于古菌基因 5′-侧翼区。另外，在古菌中还发现一种类似于 TFIIB 的蛋白质。因此，古菌的转录装置与真核生物至少具有某些相似性，提示古菌和真核生物是在共同祖先从细菌分支后进一步分化形成的。这一进化路线的推测也被古菌 rRNA 基因序列所证实，古菌 rRNA 基因序列与真核生物的相似性高于与细菌的相似性。

小结 遗传研究已经证明 TBP 突变细胞的提取物是缺陷性的，不仅不能转录 II 类基因，也不能转录 I 和 III 类基因。因此，TBP 是 3 类基因的通用必需转录因子，而且至少在某些古菌的转录中也是必需的。

TBP 相关因子 许多研究人员已经在多种生物的 TFIID 中发现了 TBP 相关因子（TBP-associated factor，TAF_{II}）。为了鉴定果蝇细胞中的 TAF_{II}，Tjian 及同事用 TBP 特异性抗体从 TFIID 粗提液中免疫共沉淀 TFIID，然后用 2.5mol/L 尿素从 TBP 抗体沉淀物中剥离 TAF，SDS-PAGE 法展示 TAF。用同样方法共鉴定出 8 种 TAF，这 8 种多肽的基因已经克隆，并用克隆基因的蛋白产物装配了 TFIID。图 11.8 显示 TFIID 中多肽的装配模式。在人和酵母的 TAF_{II} 中也进行了同样实验，3 个物种的 TAF_{II} 分子质量不同，但氨基酸序列的相似性显示了它们之间的同源性（图 11.9）。

研究人员已经发现了 TAF_{II} 的几种不同功能，其中特别受关注的两种功能是与启动子及基因特异转录因子间的相互作用。让我们来看看这两种功能的有关证据以及在每种功能中 TAF_{II} 的作用方式。

我们已经知道 TBP 在与 TATA 盒的结合中很重要，但足迹实验也表明在一些启动子中，附着在 TBP 上的 TAF_{II} 能够使 TFIID 的结合超出 TATA 盒很大范围。1994 年 Tjian 等报道，在一些启动子中，TBP 能够保护 TATA 盒周围约 20bp 的区段，而 TFIID 能够将保护区延伸到＋35 位，大大超过了转录起始位点。表明 TFIID 的 TAF_{II} 能够与启动子的起始子及下游元件作用。

为了更详细地研究这一现象，Tjian 小组用 TBP 和 TFIID 分别在体外转录含有两类启动子的 DNA。第一类（腺病毒 E1B 和 E4 启动子）含有 TATA 盒，但无起始子和下游启动子元件；第二类［腺病毒主要晚期（AdML）启动子和果蝇热激蛋白（*hsp70*）启动子］含 TATA 盒、起始子和 DPE。图 11.10 显示上述启动子的结构和体外转录实验结果。

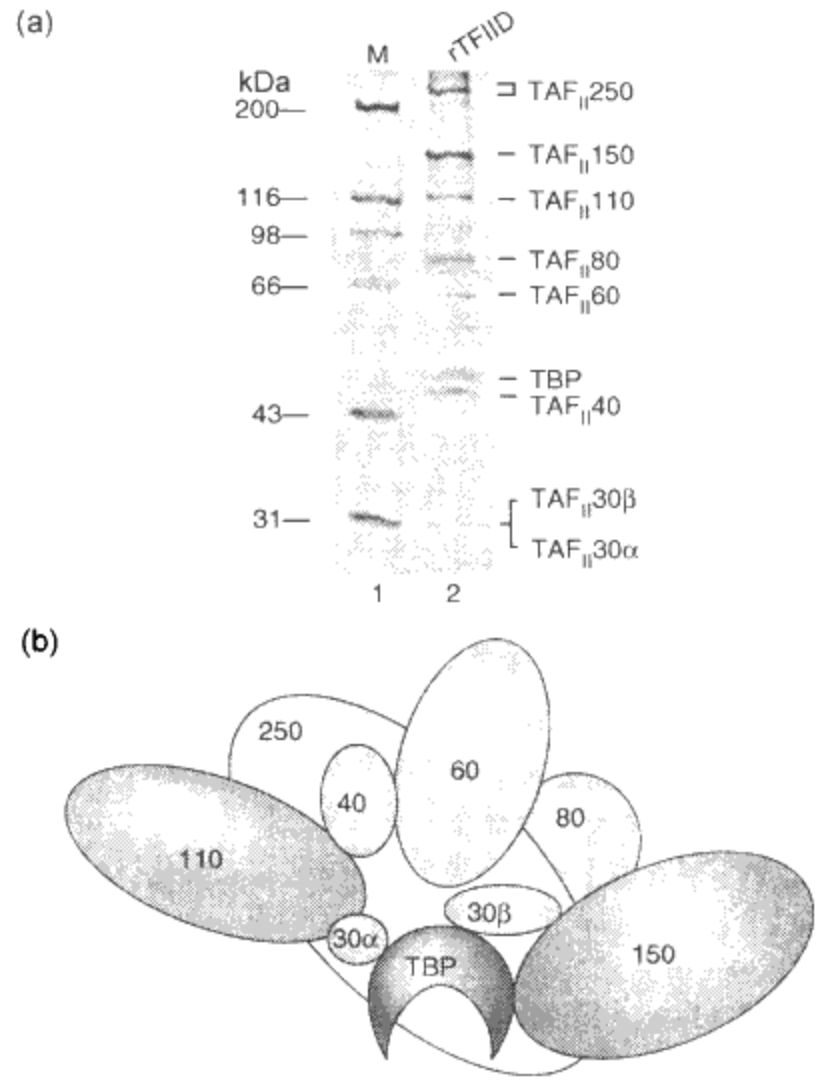


图 11.8 果蝇 TFIID 的克隆基因产物在体外的装配结构图。（a）多肽的凝胶电泳图。泳道 1 为蛋白质分子质量标准；泳道 2 为重组 TFIID（rTFIID），包含 8 个主要 TAFII。（b）推测的果蝇 TFIID 重建结构。（*Source*：Chen，J. L.，L. D. Attardi，C. P. Verrijzer，K. Yokomori，and R. Tjian，Assembly of recombinant TFIID reveals differential co-activator requirements for distinct transcriptional activators. *Cell* 79（7 Oct 1994）p. 100，f. 7a. Reprinted with permission of Elsevier Science.）

可以看到在仅含 TATA 盒的启动子上转录时，TBP 和 TFIID 的效果一致（比较泳道 1 和 2 及泳道 3 和 4），但对含有起始子和 DPE 的启动子转录时，TFIID 显示出明显的优势（比较泳道 5 和 6 及泳道 7 和 8）。因此，TFIID 中的 TAF_{II} 因子显然帮助 TBP 促进聚合酶转录带有起始子和 DPE 的启动子。

哪些 TAF_{II} 负责识别起始子和 DPE 呢？为了寻找答案，Tjian 及同事用果蝇 TFIID 与放射性标记的带有 *hsp70* 启动子的 DNA 片段进行光交联（photo-cross-linking）实验。先将溴脱氧尿苷（bromodeoxyuridine，BrdU）掺入 DNA 片段，然后让 TFIID 结合启动子，接着用紫外线照射该复合体，使蛋白质与 DNA 中的 BrdU交联。洗脱未结合的蛋白，核酸酶消化 DNA，释放标记的蛋白质并进行 SDS-PAGE。图 11.11 的泳道 1 显示两个 TAF_{II}（$TAF_{II}250$ 和 $TAF_{II}150$）因

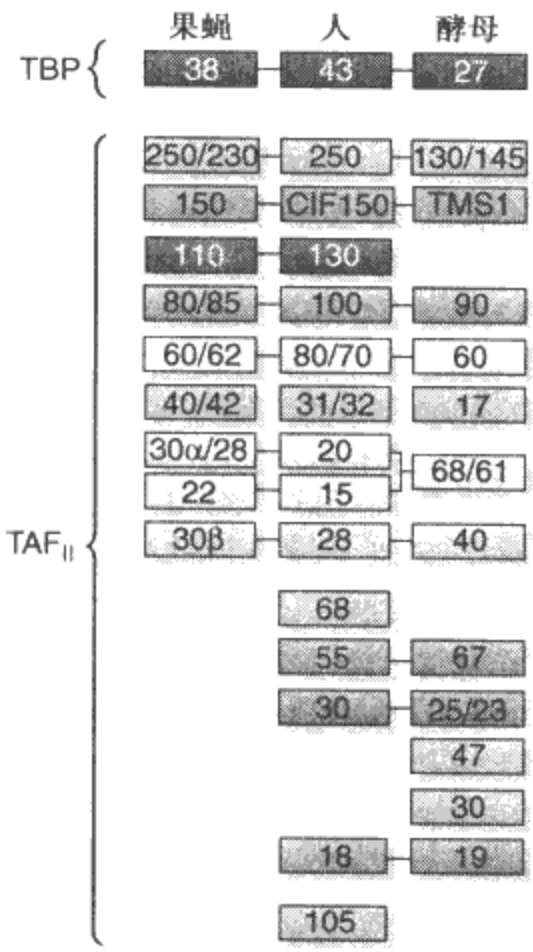


图 11.9　果蝇（*D. melanogaster*）、人（*H. sapiens*）、酵母（*S. cerevisiae*）的 TAF 之间的关系。水平线连接同源蛋白。（*Source*：Adapted from Tansey，W. P. and W. Herr，TAFs：Guilt by association? *Cell* 88：729，1997.）

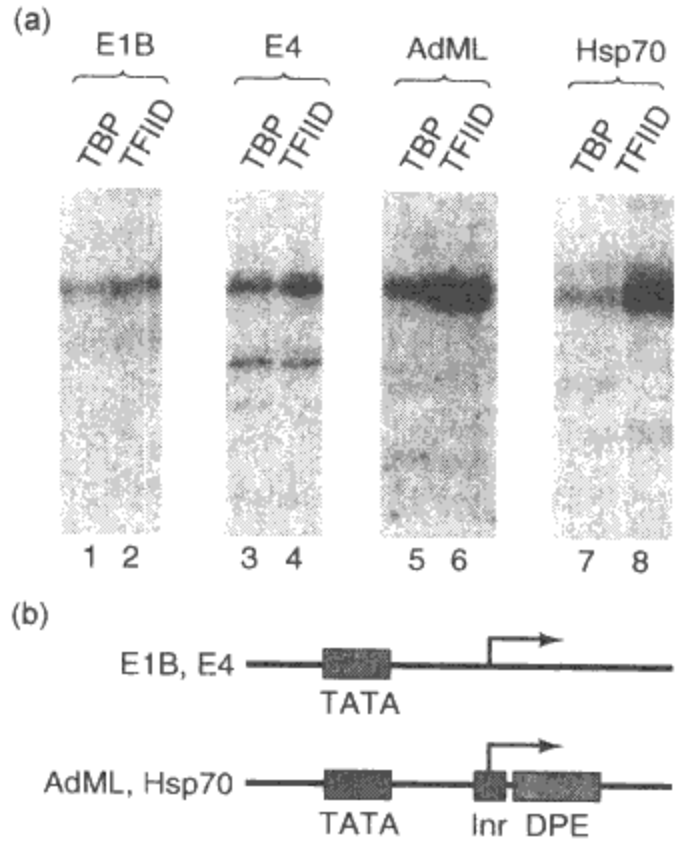


图 11.10　TBP 和 TFIID 在 4 种不同启动子上的活性。（a）实验结果。Tjian 及同事利用果蝇转录系统验证了 TBP 或 TFIID（顶部所示）在 4 种不同启动子（顶部所示）模板上的活性。两类启动子图示于（b）。第一类，腺病毒 E1B 和 E4 启动子，含有 TATA 盒。第二类，腺病毒后期主要启动子（AdML）和果蝇 *hsp70* 启动子，含有 TATA 盒及起始子和 DPE。体外转录后，利用引物延伸实验（顶部）检测 RNA 产物。放射自显影结果表明：在第一类启动子中（仅含 TATA 盒），TBP 和 TFIID 促进转录的作用相同，但在第二类启动子中（含 TATA 盒、Inr 及 DPE），TFIID 对转录的支持比 TBP 强得多。（*Source*：Verrijzer，C. P.，J.-L. Chen，K. Yokomari，and R. Tijan，Promoter recognition by TAFs. *Cell* 81（30 June 1995）p. 1116，f. 1 Reprinted with permission of Elsevier Science.）

结合到 *hsp70* 启动子而被标记；若不加 TFIID（泳道 2），则没有蛋白质被标记。Tjian 等随后重建了 TBP、TAF$_{II}$250 和 TAF$_{II}$150 三元复合物，做同样的光交联实验，泳道 3 显示该实验同样获得被标记的 TAF$_{II}$250 和 TAF$_{II}$150；泳道 4 显示 TBP 单独结合 DNA 时不能被标记。我们知道 TBP 是与含 TATA 盒的 DNA 结合的，但却不能与 BrdU 光交联，因而不被标记。这是为什么呢？很可能是因为这种光交联仅对结合到大沟的蛋白质有效，而 TBP 是结合在 DNA 小沟中的。

为了进一步确定三元复合物（TBP-TAF$_{II}$250-TAF$_{II}$150）结合的特异性，Tjian 及同事用 TBP 或三元复合物进行了 DNA 酶足迹实验。图 11.12 显示 TBP 仅在 TATA 盒中产生足迹，而三元复合物在起始子和下游元件处还产生了额外的足迹。这一点支持了两个 TAF$_{II}$结合到起始子和 DPE 上的假说。

对二元复合物（TBP-TAF$_{II}$250 或 TBP-TAF$_{II}$150）的进一步实验结果表明在识别起始子和 DPE 时，这些复合物与仅有 TBP 的情况相同。因此，两个 TAF$_{II}$似乎在增强启动子元件结合中起协同作用。进一步说明三元复合物（TBP-TAF$_{II}$250-TAF$_{II}$150）在识别由 AdML 启动子中的 TATA 盒和 TdT 启动子中的起始子组成的合成启动子时，具有与 TFIID 一样的效果。相对而言，两个二元复合物在识别启动子时的作用和单独使用 TBP 一样。这些发现支持了 TAF$_{II}$250 和 TAF$_{II}$150 在结合仅含起始子或同时含有 DPE 时均表现协同作用。

TFIID 中的 TBP 在识别大部分熟知的含有 TATA 盒的 II 类启动子过程中是相当重要的（图 11.13a），但对于缺少 TATA 盒的启动子情况如何呢？尽管这些启动子不能直接结合 TBP，但大多数启动子必须依赖 TBP 才能激活。这种看来矛盾的现象的原因在于这些缺少 TATA 盒的启动子含有其他可以使 TBP 结合的元件。这些元件可以是起始子和 DPE，使得 TAF$_{II}$250 和 TAF$_{II}$150 能够结合上去，从而保证整个 TFIID 结合到启动子上（图 11.13b）。

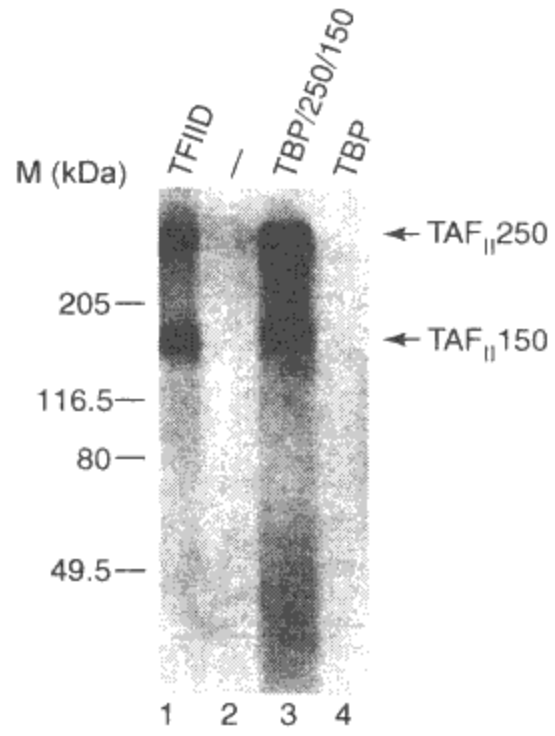


图 11.11　鉴定结合在 *hsp70* 启动子上的 TAF_{II}。Tjian 及同事将 TFIID 光-交联到含有 *hsp70* 启动子并用^{32}P 标记的模板上。先将 TFIID 结合到标记模板上，该模板已被光敏核苷酸 BrdU 取代。接着用紫外光照射 TFIID-DNA 复合物，在 DNA 与紧密接触 DNA 大沟的蛋白质之间形成共价键。随后用核酸酶消化 DNA，并进行 SDS-PAGE。泳道 1 为加入 TFIID 的自显影结果。TAF_{II}250 和 TAF_{II}150 被标记，表明这两种蛋白质与标记的 DNA 大沟有紧密联系；泳道 2 为无 TFIID 的对照；泳道 3 是加入三元复合物 TBP-TAF_{II}250-TAF_{II}150 的结果，两个 TAF_{II} 被标记，再次表明它们与 DNA 结合；泳道 4 是加入 TBP 的结果，因为不与 DNA 大沟结合而未被标记，与预期结果一致。（*Source*：Verrijzer，C. P.，J.-L. Chen，K. Yokomari，and R. Tijan，*Cell* 81（30 June 1995）p. 1117，f. 2a. Reprinted with permission of Elsevier Science.）

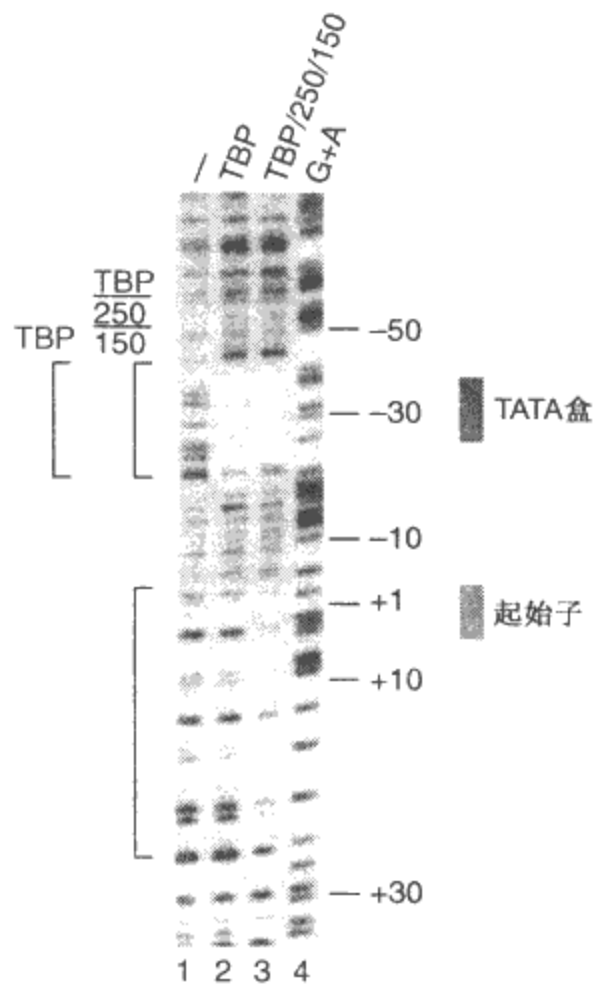


图 11.12　TBP 和三价复合物（TBP、TAF_{II}250 和 TAF_{II}150）在 *hsp70* 启动子上的 DNase I 足迹。泳道 1，无蛋白质。泳道 2，TBP。泳道 3，三元复合物。在泳道 2 和 3 都加入了 TFIIA 以稳定 DNA-蛋白质复合体，但其他实验表明 TFIIA 的加入不影响足迹的范围。泳道 4 是 G+A 测序泳道作对照。左边括号标出了由 TBP 和三元复合物产生的足迹范围。右边的条框指出 TATA 盒和起始子的位置。（*Source*：Verrijzer，C. P.，J.-L. Chen，K. Yokomori，and R. Tijan，*Cell* 81（30 June 1995）p. 1116，f. 2c. Reprinted with permission of Elsevier Science.）

结合基因特异转录因子的上游元件也可起到上述作用，它们依次与一个或多个 TAF_{II} 相互作用，将 TFIID 锚定到启动子上。例如激活因子 Sp1 结合到上游元件（GC 框）上，并至少与一个 TAF_{II}（果蝇中的 TAF_{II}110，人类的 TAF_{II}130）发生作用，这一中介作用显然有助于 TFIID 结合到启动子上（图 11.13c）。

TAF_{II}的第二个主要活性是参与激活因子介导的转录激活，其中部分内容将在第 12 章介绍。1990 年，Tjian 及同事证明 TFIID 足以支持由 Sp1 介导的激活，但 TBP 却不能。这些结果表明 TFIID 中的某些因子是其与上游作用因子（upstream acting factor）如 Sp1，相互作用所必需的，而这些因子在 TBP 中没有。从定义上讲，这些因子是 TAF_{IIs}，有时也称为**辅激活因子**（coactivator）。

由此可知，将 TBP 与某些 TAF_{IIs}混合能产生参与特定启动子转录的复合物。例如，TBP-TAF_{II}250-TAF_{II}150 三价复合物对 TATA 盒和起始子组成的启动子识别几乎与完整 TFIID 一样好。Tjian 及同事用类似技术研究了哪个 TAF_{II} 参与 Sp1 引起的转录激活。他们发现在果蝇和人提取物中只有当 TAF_{II}110 存在时 Sp1 才能激活转录。因此，TBP、TAF_{II}250 和 TAF_{II}150 虽足以维持本底水平的转录，但不支持由 Sp1 激活的转录。在以上三因子中添加 TAF_{II}110 就允许 Sp1 激活转录。

Tjian 及同事还证明 Sp1 是直接结合在 TAF_{II}110 上而不是在 TAF_{II}250 和 TAF_{II}150 上的。他们制备了一个含有 GC 盒和 Sp1 的亲和柱验证 Sp1 结合 3 个 TAF_{II}的能力。正如所料，仅有 TAF_{II}110 被保留。

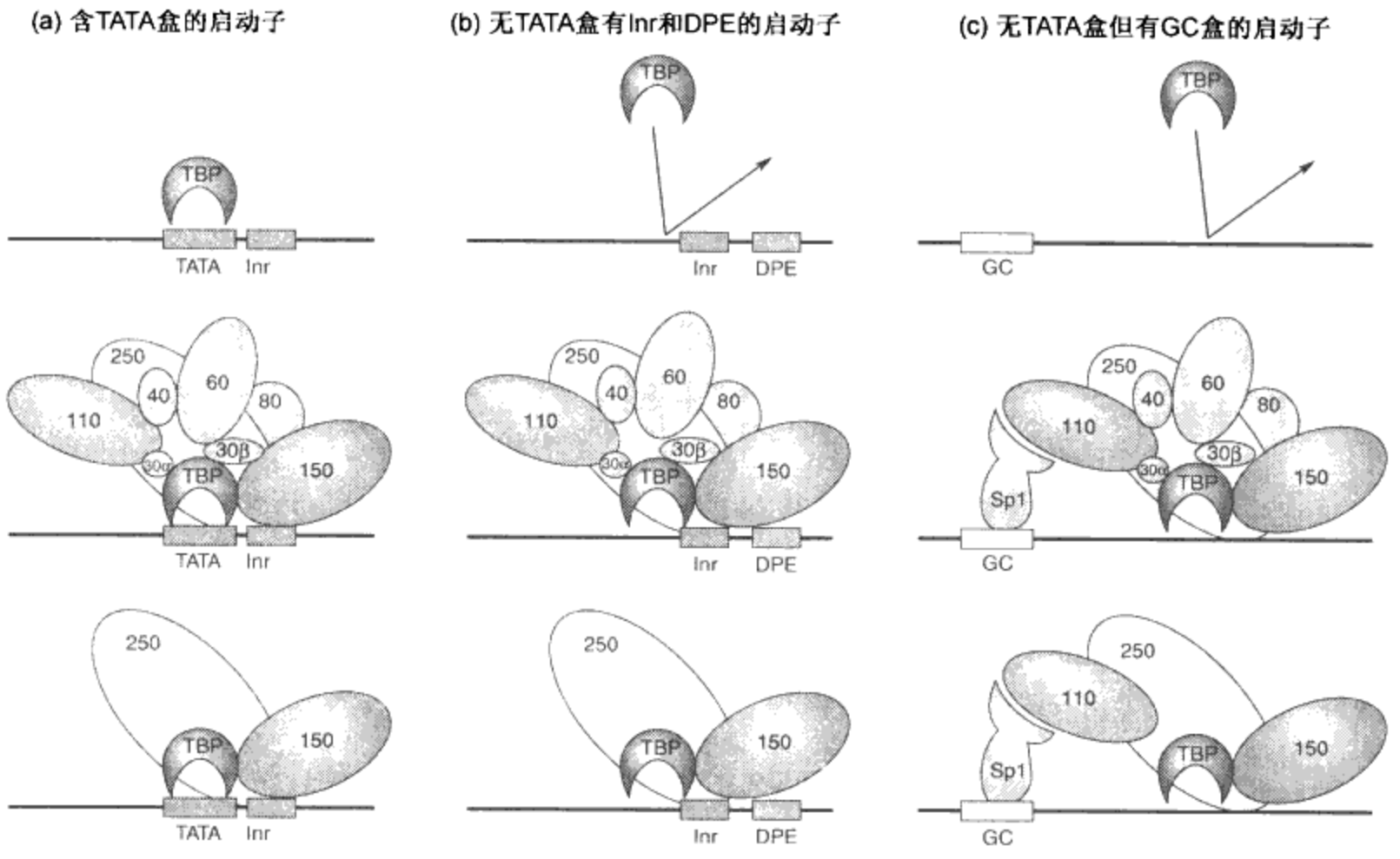


图 11.13　TBP 与有或无 TATA 盒的启动子作用模型。（a）含 TATA 盒的启动子。TBP 可以独立结合到启动子上的 TATA 盒上（顶部），也可以和 TFIID 的所有 TAF_{II}一同结合（中部），还可以与部分 TAF_{II}一同结合（底部）。（b）有起始子和 DPE 的无 TATA 盒启动子。TBP 不能独立结合到不含 TATA 盒的启动子上（顶部）。通过 $TAF_{II}250$（黄色）和 $TAF_{II}150$（褐色，中部）之间的相作用，整个 TFIID 可以结合无 TATA 盒的启动子。$TAF_{II}250$ 和 $TAF_{II}150$ 足以将 TBP 绑定到起始子和 DPE 上（底部）。（c）无 TATA 盒但有 GC 盒的启动子。TBP 不能独立结合到该启动子上（顶部）。通过与结合了 GC 盒的 Sp1 相互作用，整个 TFIID 可以与启动子结合（中部）。$TAF_{II}250$、$TAF_{II}150$ 和 $TAF_{II}110$ 足以锚定 TBP 至结合了 GC 盒的 Sp1 上。（*Source*：Adapted from Goodrich, J. A., G. Cutter, and R. Tjian, Contacts in context: Promoter specificity and macromolecular interactions is transcription. *Cell* 84: 826, 1996.）

用同样的策略，他们还证明了另一个激活因子 NTF-1 与 $TAF_{II}150$ 结合。NTF-1 需要 $TAF_{II}250$ 和 $TAF_{II}150$ 或 $TAF_{II}250$ 和 $TAF_{II}60$ 才能激活体外转录。因此，不同激活因子与不同 TAF_{II}组合相作用可增强转录，而所有组合中都有 $TAF_{II}250$。表明 $TAF_{II}250$ 作为装配因子，其他 TAF 围绕它而聚合。这些发现与图 11.14 的模型一致，每个激活因子与一组特定 TAF_{II}作用，所以完整 TFIID 能同时与几个激活因子相互作用，放大它们的效果，并产生很强的转录增强作用。

除了具有与启动子元件和激活因子作用的能力外，TAF_{II}还具有酶活性。其中研究最清楚的是 $TAF_{II}250$，已知它有两种酶活性。一种是组蛋白乙酰转移酶（histone acetyltransferase，HAT）活性，可将乙酰基转移到组蛋白的赖氨酸残基上。这种乙酰化作用一般是激活转录事件；它还是一种蛋白激酶，可以将自身和 TFIIF（还有 TFIIA 和 TFIIE，尽管在较低程度上）磷酸化。这些磷酸化作用可调节前起始复合物的装配效率。

虽然前面的研究显示在体外的前起始复合物形成中不需要 TFIIA，但它对 TBP（或 TFIID）结合启动子却是必需的。许多研究均得出此结论，而其中一个强有力的证据是：编码酵母 TFIIA 两个亚基的任何一个基因突变都是致死的。

TFIIA 不仅稳定了 TBP-TATA 盒的结合，也通过一种抗阻抑（antirepression）机制促进 TFIID-启动子结合。过程如下：TFIID 未结合启动子时，TBP 的 DNA 结合表面被 $TAF_{II}250$ 的 N 端域所覆盖，因而抑制了 TFIID 对启动子的结合。而 TFIIA 可干扰 $TAF_{II}250$ N 端域对 TBP 的 DNA 结合表面的作用，释放 TBP，使其与启动子结合。

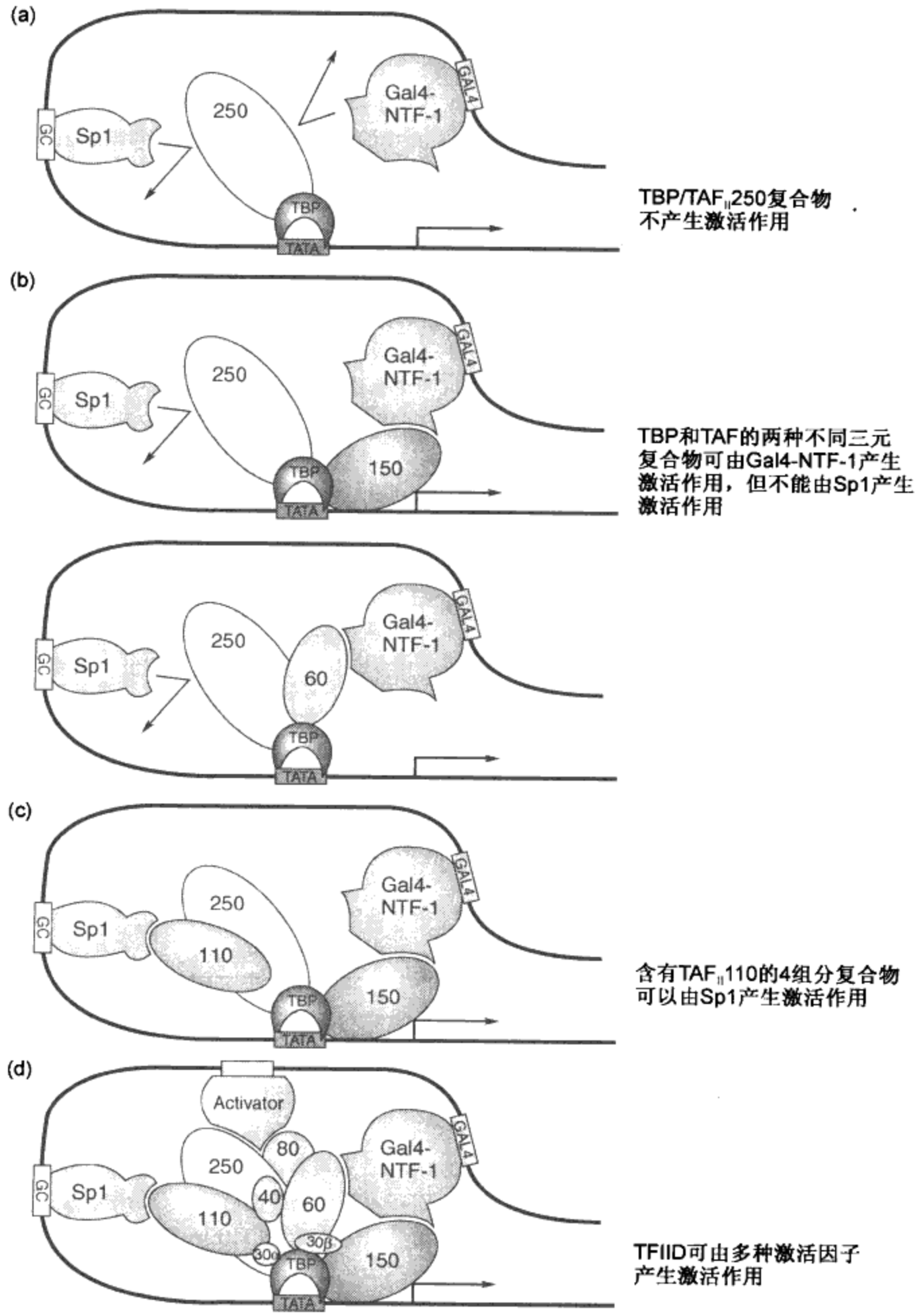


图 11.14　激活因子增强转录的作用模型。(a) $TAF_{II}250$ 既不与 Sp1 也不与 Gal4-NTF-1（带有 NTF-1 转录激活域的杂合激活因子）作用，因此无激活作用。(b) Gal4-NTF-1 与 $TAF_{II}150$ 或 $TAF_{II}60$ 作用，激活转录。而 Sp1 与这两种 TAF 以及 $TAF_{II}250$ 均不能作用，转录不能激活。(c) Gal4-NTF-1 与 $TAF_{II}150$、Sp1 与 $TAF_{II}110$ 作用，两个因子均可激活转录。(d) 完整 TFIID 包含所有 TAF_{II}，能对各种激活因子做出响应，这里给出的激活因子为 Sp1、Gal4-NTF-1 和通用激活因子（绿色，位于图顶部）（*Source*：Adapted from Chen, J. L., L. D. Attardi, C. P. Verrijzer, K. Yokomori, and R. Tjian, Assembly of recombinant TFIID reveals differential coactivator requirements for distinct transcriptional activators. *Cell* 79: 101, 1994.）

小结 TFIID 至少包括 8 种 TAF_{II}，除了 TBP 外，大多数 TAF_{II}在真核生物进化中是保守的。TAF_{II}行使多种功能，两个显著的功能是与核心启动子元件和激活因子的相互作用。TAF_{II}250 和 TAF_{II}150 有助于 TFIID 结合启动子的起始子和 DPE，使 TBP 与无 TATA 盒但有起始子和 DPE 的启动子结合。TAF_{II}250 和 TAF_{II}110 有助于 TFIID 与结合在转录起始位点上游 GC 盒上的 Sp1 相互作用，使 TBP 可以结合无 TATA 盒但有 GC 盒的启动子。与各种激活因子的作用，显然需要不同 TAF_{II}的组合，至少在高等真核生物是这样。TAF_{II}250 还具有两种酶活性，即组蛋白乙酰转移酶活性和蛋白激酶活性。

TAF 和 TBP 因子通用性的例外 酵母遗传学研究对图 11.14 模型的通用性提出了质疑。Michael Green 和 Kevin Struhl 及同事发现酵母 TAF 基因的突变体是致死的，但却不影响转录活性，至少在第一批被研究的基因中是这样的。例如，Green 及同事在基因中引入了温度敏感型突变子，这些基因编码酵母 $yTAF_{II}$145，与果蝇和人的 TAF_{II}250 同源。他们发现当超过允许的温度范围时，$yTAF_{II}$145 的浓度迅速下降，而且至少还有其他两个 $yTAF_{II}$浓度也降低。$yTAF_{II}$145 的失活明显破坏了 TFIID，进而引起其他 $yTAF_{II}$的降解。尽管这些 $yTAF_{II}$失活了，但由各种激活因子激活的 5 种不同基因在体内的转录效率并未受到不适温度的影响。研究者在 $yTAF_{II}$30 基因敲除突变体中也获得了同样结果。与此相反，当编码 TBP 或 RNA 聚合酶亚基的基因突变时，所有转录很快就停止了。

Green、Richard Young 及同事基于前人的研究，在全基因组范围内分析了两个 TAF_{II}基因及其他几个酵母基因突变所产生的效应。在 TAF_{II}145（与哺乳动物 TAF_{II}250 同源）和 TAF_{II}17（与哺乳动物 TAF_{II}31/32 同源）中引入温度敏感型突变，然后升高温度使 TAF 失活或降低温度至 TAF 有活性，用高密度基因芯片确定了 5460 个酵母基因的表达。这些芯片包含了每个基因特有的寡聚核苷酸序列。酵母总 RNA 可与这些序列杂交，与每个寡核苷酸杂交的程度可以衡量相关基因表达的程度。此实验分别比较了低温和高温下的 RNA 与寡核苷酸杂交的情况，并将结果与 RNA 聚合酶 II 最大亚基（Rpb1）的温度敏感型突变效果进行了比较。因为 Rpb1 突变可阻止所有 II 类基因的转录，为比较其他基因突变效应提供了基准。

表 11.1 显示了以上分析结果。只有 16%的酵母基因，像依赖于 Rpb1 一样，依赖 TAF_{II}145，表明 TAF_{II}145 在 16%的酵母基因转录中是必需的。如果 TAF 是 TFIID 的基本组成部分，而 TFIID 又是所有 II 类基因前起始复合物的基本组分，那么这就不是所期望的结果。确实，TAF_{II}145 和 TBP 被认为是 TFIID 中的重要作用因子，可以在 TAFIID 中装配其他所有 TAF，但是全基因组表达分析显然并不支持这一观点。事实上，在高等生物中 TAF_{II}145 和它的同源物只在部分基因的前起始复合物形成过程中是必需的。在酵母中，这些基因可能调节细胞周期发育。

酵母其他 TAF（TAF_{II}17）的突变产生了更显著的效应。在所研究的酵母基因中，有 67%都依赖于 TAF_{II}17（像依赖 Rpb1 一样）。但这并不意味着 TFIID 就是所有这些基因转录所必需的，因为 TAF_{II}17 同时还是转录适配复合物（transcription adapter complex）也叫做 **SAGA** 的一部分（SAGA 命名源于它包含的三类蛋白质 SPT、ADA 及 GCN5）。与 TFIID 一样，SAGA 包含 TBP、若干 TAF 及 HAT 活性，似乎具有介导某些转录激活因子的效应。所以突变 TAF_{II}17 的效应可能在于它在 SAGA 中的作用或其他还未发现的蛋白质复合物中的作用，而不是在 TFIID 中的作用。

表 11.1 酵母需求转录因子的全基因组分析

通用转录因子（亚基）	依赖于亚基功能的基因比例/%
TFIID（TAF_{II}145）	16
TFIID（TAF_{II}17）	67
TFIIE（Tfa1）	54
TFIIH（Kin28）	87

不仅某些 TAF 不是所有转录所必需的，连 TFIID 中 TAF 的组成看起来也是异源的（heterogenous）。例如，TAF_{II}30 仅在人的部分 TFIID 中发现，且它和雌激素的响应有关。

更令人惊奇的是，甚至 TBP 也不是普遍存在于高等真核生物的前起始复合物中。最显著的例子是果蝇 TBP 的替代物 **TRF1**（TBP-相关因子 1，TBP-related factor 1）。该蛋白质在发育的神经组织中表达，与 TBP 一样结合在 TFIIA 和 TFIIB 上促进转录，还有自身的一组 TRF 相关因子（TRF-associated factor），称为 **nTAF**（神经系统的 TAF）。2000 年，Michael Holmes 和 Robert Tjian 用引物延伸法分别在体内和体外证明了 TRF1 能够促进果蝇 *tudor* 基因的转录，而且分析还表明该基因有两个不同的启动子：一个是有 TATA 盒的下游启动子，被含有 TBP 的复合物识别；第二个位于第一个启动子上游 77bp 处，有 TC 盒，被含有 TRF1 的复合物识别（图 11.15）。TC 盒从－22bp 到－33bp，与转录起始有关，在非模板链的序

列为 AITGCTTTTCTT。在 DNase 足迹实验中，TC 盒被 TRF1、TFIIA 和 TFIIB 复合物所保护。但是这 3 个蛋白质单独都不能在该区域产生足迹，TBP 或 TBP+TFIIA 和 TFIIB 也不能产生足迹。

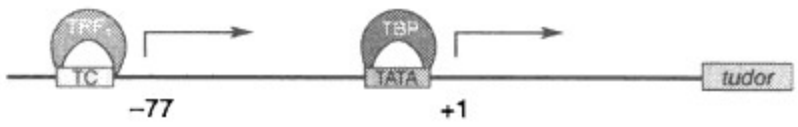


图 11.15　果蝇 *tudor* 基因调控区。该基因有两个相距约 77bp 的启动子。下游启动子有 TATA 盒，吸引基于 TBP 的前起始复合物。上游启动子有 TC 盒，吸引基于 TRF1 的前起始复合物。

因此，TRF 可能是 TBP 的细胞类型特异性变体。TBP 和 TRF 替代物的存在产生了一种可能性，即高等真核生物基因的表达调控可能部分是取决于合适的 TBP、TAF 以及激活因子蛋白的可用性。激活因子将在第 12 章介绍。的确，两个不同 TBP 识别 *tudor* 基因的两个不同启动子，使人联想到原核生物 RNA 聚合酶的不同 σ 因子对同一基因的两个不同启动子的识别（第 8 章）。

实际上，TRF 好像是为果蝇所特有。但是目前所研究的所有多细胞动物中有另一种 **TBP 类似因子**（TBP-like factor，**TLF**）。与 TBP 不同，TLF 缺少插入 TATA 盒碱基对并帮助启动子 DNA 弯曲的成对的苯丙氨酸。因此，TLF 不结合 TATA 盒，而是调控其他无 TATA 盒启动子的转录。

在前起始复合物形成中，TBP 的核心作用受到了质疑。研究发现无 TBP 的 TAF_{II} 复合物 TFTC（TBP-free TAF_{II}-containing complex）无需 TFIID 和 TBP 协助就可自发形成前起始复合物。Patrick Schultz 及同事基于结构研究提出了 TFTC 替代 TFIID 的想法。电镜和数码图像分析发现两者在三维结构上非常相似。图 11.16 显示了两个蛋白质复合物在 3 个不同方向的三维模型。最显著的特征是两者都有一个很大的沟，足以结合双链 DNA。实际上，它们可以像夹子一样夹住 DNA。主要差异在于 TFTC 顶端因有结构域 5 而有一个突起，TFIID 没有结构域 5 和突起。

在第 10 章已经学习过果蝇的许多启动子都没有 TATA 盒，而有 DPE，并且通常与起始子偶联。我们还学过 DPE 通过其一个或几个 TAF 与 TFIID 结合。2000 年，James Kadonage 及同事在果蝇中又发现了一个因子（dNC2）与其他生物中熟知的**负辅因子 NC2**（negative cofactor 2）或 Drl-Drap1 因子同源。为简便起见，将所有此类因子均称作 NC2。他们还发现了一个有趣的现象：NC2 可以区别含 TATA 盒的启动子和含 DPE 的启动子。实际上，NC2 促进含 DPE 的启动子转录，抑制含 TATA 盒的启动子转录。因此，NC2 可能是基因调控的焦点。

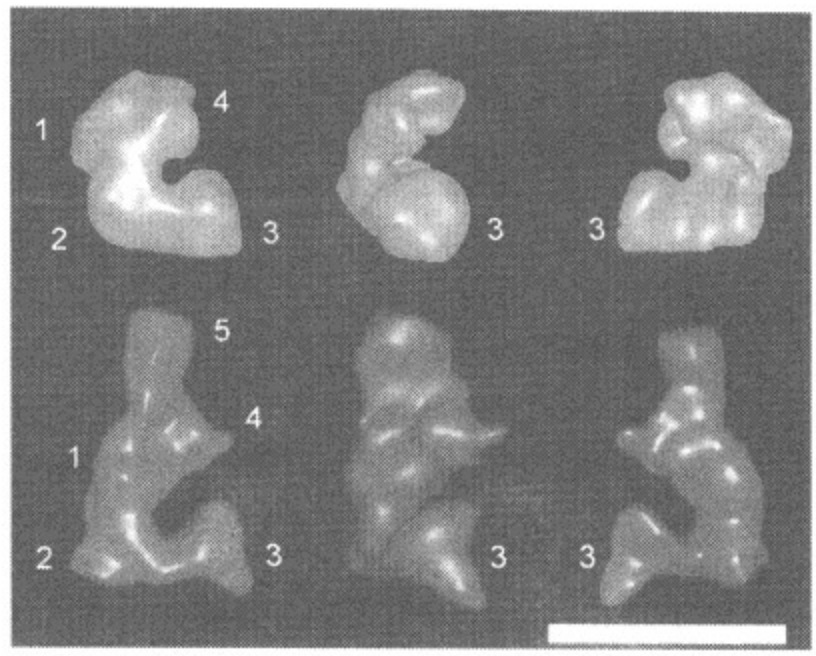


图 11.16　TFIID 和 TFTC 的三维模型。Schultz 及同事制作了 TFIID 和 TFTC 的负染电镜图（方法见第 19 章），数字化合成图像。然后在电子显微镜下旋转不同的角度，收集两种蛋白质的三维图像信息。图上显示分析所得模型，TFIID（绿色），TFTC（蓝色）。（*Source*：Brand，M.，C. Leurent，V. Mallouh，L. Tora，and P. Schuttz，Three-dimensional structures of the TAF_{II}-containing complexes TFDIID and TFTC. *Science* 286（10 Dec 1999）f. 3，p. 2152. Copyright © AAAS.）

2001 年，Stephen Burley 及同事解析了 NC2-TATA 盒-TBP 复合体的晶体结构，揭示了 NC2 如何抑制 TATA 盒启动子的转录。NC2 结合在被马鞍形 TBP 折弯的 DNA 的下侧，一旦 NC2 结合到启动子上，它的一个 α 螺旋就会阻止 TFIIB 与复合物结合，同时，NC2 的另一部分会干扰 TFIIA 的结合。而没有 TFIIA 和 TFIIB，前起始复合物不能形成，转录也就无法起始。

小结　在 II 类基因转录中，TAF 并不是通用的。$TAF_{II}145$（与人的 $TAF_{II}250$ 同源的酵母基因）在酵母大部分 II 类基因的转录中也不通用，甚至 TBP 也不是普适的。在高等真核生物中，某些启动子对 TBP 的替代物如 TRF1 产生应答。某些启动子可以被不含 TBP 的 TAF_{II} 复合物（TFTC）而不是 TFIID 所激活。通用转录因子 NC2 促进含 DPE 的启动子但抑制 TATA 启动子的转录。

TFIIB 的结构和功能

Danny Reinberg 及同事克隆并表达了人 TFIIB 基因。在所有体外实验中，包括对激活因子如 Sp1 的应答，克隆 TFIIB 基因的产物都可以代替天然人

TFIIB蛋白。表明TFIIB是一个单亚基因子（M_r=35kDa），无需TAF类似的辅助多肽。正如已经了解的那样，在体外TFIIB是继TFIID和TFIIA之后结合到前起始复合物上的第三个或第二个（如果TFIIA还未结合）转录因子。这是RNA聚合酶结合所必需的，因为TFIIF-聚合酶复合物需结合在DAB复合物而不是DA复合物上。

在前起始复合物装配过程中，TFIIB结合的顺序位于TFIID和TFIIF/RNA聚合酶II之间，表明TFIIB有两个结构域，分别与以上两个蛋白质结合。事实上，TFIIB确实有两个不同的域：N端域（$TFIIB_N$）和C端域（$TFIIB_C$）。2004年Kornberg及同事对TFIIB的结构研究发现，这两个结构域的确在连接TATA盒处的TFIID和RNA聚合酶II之间起桥梁作用，使聚合酶的活性中心位于TATA盒下游25～30bp处，正好是转录起始处。特别是该研究还揭示了TBP通过弯曲TATA盒处的DNA，使DNA包绕$TFIIB_C$，而$TFIIB_N$则与聚合酶的一个位点结合，将其正确地置于转录起始位点上。

Kornberg及同事从酿酒酵母中获得了RNA聚合酶II-TFIIB复合物的晶体。图11.17显示该复合物晶体结构的两种视图，并加入了早期研究所推测的TBP和启动子在晶体中的位置，可以看到在复合物中TFIIB的两个结构域$TFIIB_C$和$TFIIB_N$。$TFIIB_C$（紫红色）与TBP和TATA盒处的DNA作用，而被TBP弯曲的DNA（在TATA盒处）的确包裹在$TFIIB_C$和聚合酶周围。弯曲的DNA直接伸向位于聚合酶活性中心附近的$TFIIB_N$。

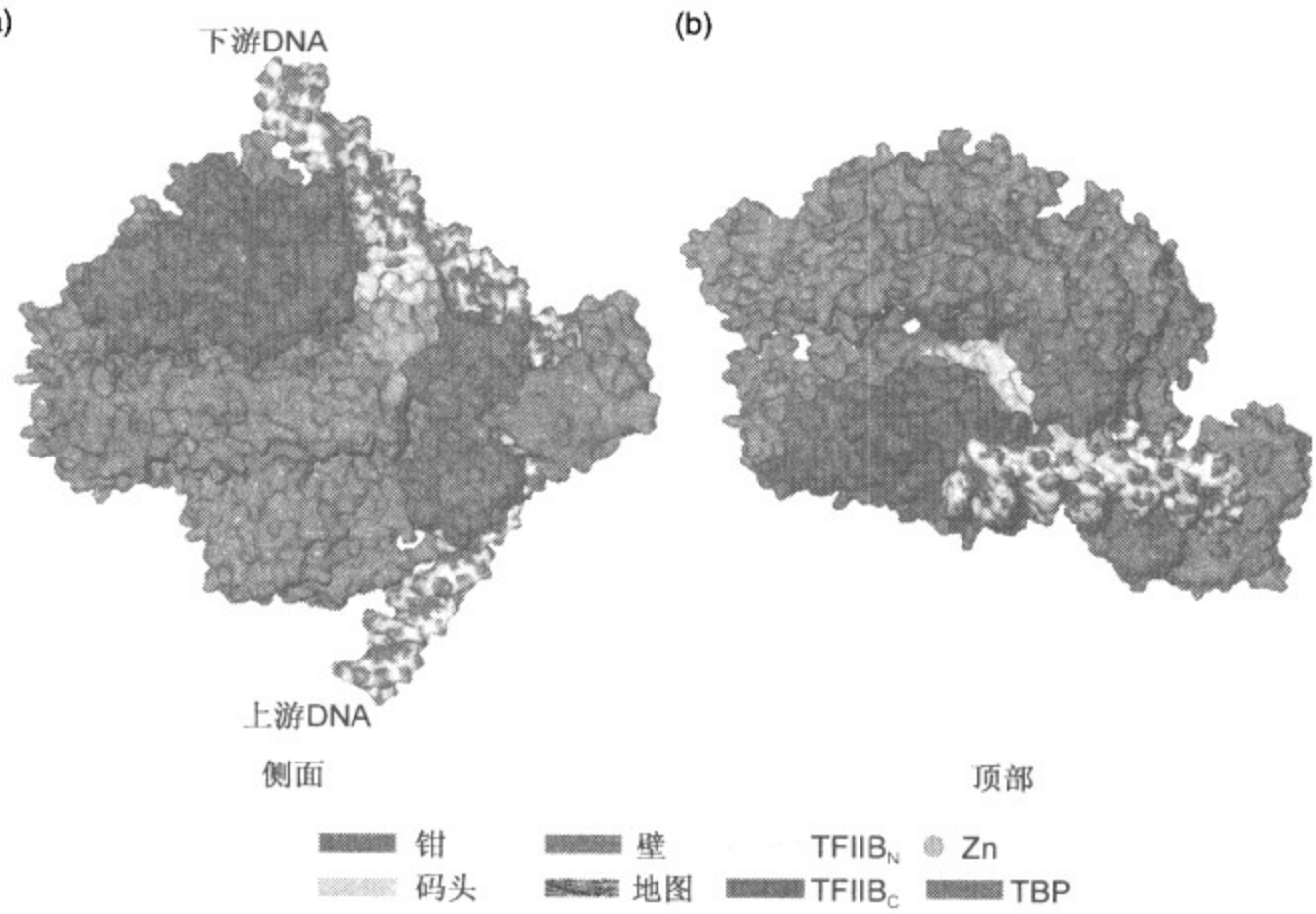


图11.17 TFIIB-TBP-聚合酶II-DNA复合体的结构模型。（a）和（b）显示Kornberg及同事从两个独立复合物的结构推测的结构：一个是$TFIIB_C$-TBP-TATA盒DNA，另一个是RNA聚合酶II-TFIIB，显示两种不同视角的结构。底部的颜色图标用以区分蛋白质及结构域，聚合酶的其他区域表示为灰色，弯曲的TATA盒DNA及其20bp的B型DNA延伸部分，以红白和蓝表示。（*Source*：（*a-b*）Reprinted with permission from *Science*, Vol. 303, David A, Bushnell, Kenneth D. Westover, Ralph E. Davis, Roger D. Kornberg, "Structural Basis of Transcription: An RNA Polymerase II-TFIIB Cocrystal at 4.5 Angstroms" Fig. 3 c&d, p. 986, Copyright 2004, AAAS.）

早期研究已表明$TFIIB_N$上的突变能改变转录起始位点，而目前研究为其提供了理论依据。特别是已知在62～66位残基处的突变可改变起始位点。这些氨基酸位于$TFIIB_N$的指形域一侧，与DNA模板链上的－6～－8位碱基（起始位点为＋1）接触（图11.18的左上部）。而且，指尖靠近聚合酶的活性中心，并位于启动子中的起始子（围绕起始位点）的附近（第10章）。

人类TFIIB的指尖结构含有两个能够与起始子DNA很好结合的碱性残基（赖氨酸），使转录起始定位于此处。但在酵母TFIIB中，这两个碱性氨基酸被两个酸性氨基酸取代，且酵母启动子中没有起

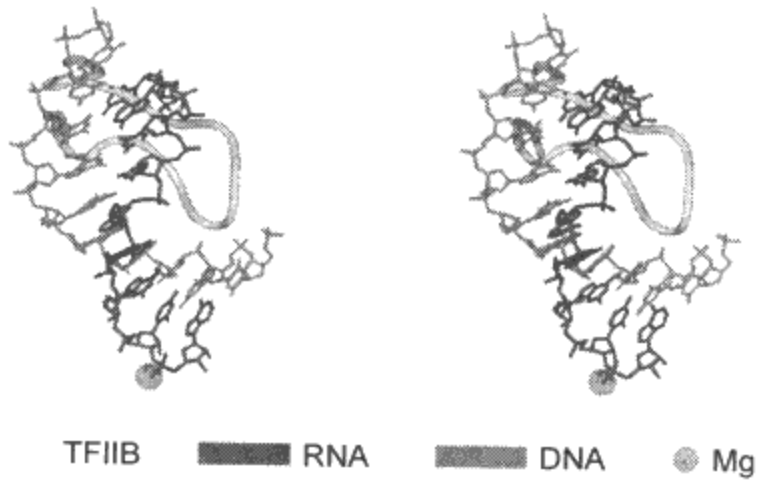


图 11.18　$TFIIB_N$ 的 B 指结构、DNA 模板链及 RNA 产物间相互作用的立体视图。结构中的各元件用底部所示不同颜色区分。（*Source*：Reprinted with permission from *Science*，Vol 303，David A. Bushnell，Kenneth D. Westover，Ralph E. Davis，Roger D. Kornberg，“Structural Basis of Transcription：An RNA Polymerase II-TFIIB Cocrystal at 4.5 Angstroms” Fig. 4，p. 987. Copyright 2004，AAAS.）

始子序列。这些考虑可能有助于解释为什么人类前起始复合物能够在 TATA 盒下游 25～30bp 处成功起始转录，而酵母的转录起始位点则有很大变动范围（在 TATA 盒下游 40～120bp）。

Kornberg 及同事总结认为 TFIIB 在定位转录起始位点过程中起双重作用。首先是粗定位，通过 $TFIIB_C$ 域在 TATA 盒处结合 TBP，以及 $TFIIB_N$ 域的指结构和相邻的锌带结构与聚合酶结合而实现。在大多数真核生物中，这可使聚合酶置于 TATA 盒下游 25～30bp 的起始位点上。然后，当 DNA 解旋时通过 $TFIIB_N$ 域的指结构与紧邻起始子上游的 DNA 作用实现精确定位。注意 TFIIB 不仅决定转录的起始位点，也决定转录的方向。这是因为它与启动子的非对称性结合，以 C 端域结合启动子上游，N 端域结合启动子下游，由此造成了前起始复合物的不对称性，进而确立了转录的方向。

以上的模型虽好，但却是几个独立的部分结构拼接在一起的，难免会猜想它与完整前起始复合物结构有多接近？为探索这一问题，Hung-Ta Chen 和 Steven Hahn 利用光交联和羟基自由基切割分析（hydroxyl radical cleavage analysis）相结合的方法对酵母 TFIIB 的各结构域与酵母 RNA 聚合酶 II 各域间的作用进行了绘图。

羟基自由基切割分析策略如下：通过定点突变（第 5 章）为蛋白质引入半胱氨酸残基，然后在每个半胱氨酸上附加一个 Fe-EDTA 复合物 Fe-BABE，它可产生羟基自由基，在 15Å 内切割蛋白质链。切割后的蛋白质片段由凝胶电泳和 Western 印迹展示。此方法可鉴定第二个蛋白质的任何区域，只要该区域位于第一个蛋白质上给定半胱氨酸的 15Å 范围内。

在第一个实验中，Chen 和 Hahn 将 TFIIB 的指形域和接头区的多个氨基酸变成了半胱氨酸，然后连接 Fe-BABE。用修饰的 TFIIB 分子装配前起始复合物，在 TFIIB 的指域和接头区激发羟基自由基以切割靠近半胱氨酸的蛋白质。为方便 Western 印迹，在 Rpb1 或 Rpb2 的末端连接了抗原决定簇（FLAG），因此可用抗 FLAG 的抗体检测 Western 印迹。图 11.19a～c 显示用抗 FLAG 抗体检测的 Western 印迹结果，FLAG 加于 Rpb2 的 N 端或 C 端或 Rpb1 的 C 端。由羟基自由基切割产生的新条带用方括号标出，在不替换半胱氨酸的野生型（wt）泳道或无 Fe-BABE（－）泳道未发现新条带。

这些条带含有已知长度的蛋白质片段，并包含蛋白质的 N 端或 C 端，利用抗 FLAG 抗体可以检测附着于蛋白质末端的 FLAG 抗原决定簇，由此可将切割位点绘制到已知的蛋白质晶体结构的相应位置。图 11.19d 是另一个类似实验的结果，其中未使用 FLAG 抗原决定簇，而是用 Rpb1 的 N 端 200 个氨基酸残基中的天然表位制备的抗体。

基于此信息，Chen 和 Hahn 绘制了 Rpb1 和 Rpb2 上与带有 Fe-BABE 的半胱氨酸紧密接触的部分。图 11.19e 和 f 分别是 TFIIB 变体（半胱氨酸被引入指域和接头区）所形成的切割图，深蓝和浅蓝分别表示强切割和中或弱切割（译者注：原文为浅蓝色表示强切割），是 Rpb1 和 Rpb2 上与 TFIIB 的指域和接头区紧密接触的区域。两个图的相似性揭示，在前起始复合物中 TFIIB 的指域和接头区很靠近，而且正如预测的那样：TFIIB 的这一部分（$TFIIB_N$）确实与 RNA 聚合酶 II 接触，具体位点是聚合酶靠近活性中心的突起、壁、钳子和叉。

在光交联实验中，Chen 和 Hahn 将 ^{125}I-标记的光交联剂（称为 PEAS）连接到与羟基自由基切割实验相同的 TFIIB 半胱氨酸变体的半胱氨酸残基上。用这些变体 TFIIB 装配前起始复合物，辐射复合物使之形成共价交联，然后通过 SDS-PAGE 观察交联并通过放射自显影检测 ^{125}I 标签。正如所预测，TFIIB 的指域和接头区与 RNA 聚合酶 II 发生了交联。预料之外的是 TFIIB 的指域和接头区也与 TFIIF 的最大亚基发生了交联，使其处于靠近聚合酶 II 活性中心的位置。

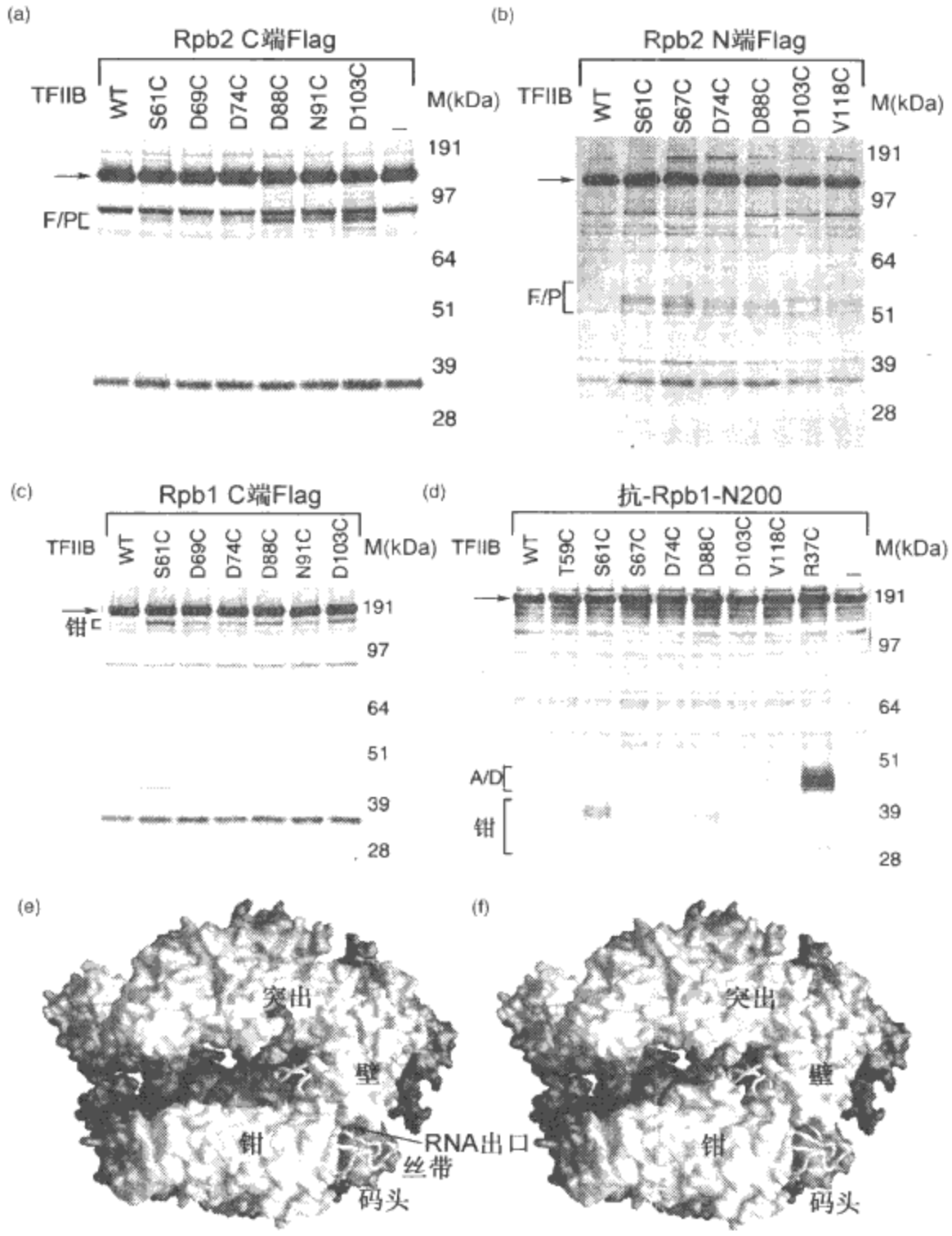


图 11.19　酵母前起始复合物中 TFIIB 与 RNA 聚合酶 II 间接触图。（a～d）Chen 和 Hahn 将 TFIIB 的指域和接头区的多个氨基酸（氨基酸位置标在泳道顶部）变为半胱氨酸，然后连接 Fe-BABE（羟基自由基发生剂）。用修饰后的 TFIIB 和 Rpb2 的 C 端（a）或 N 端（b）或 Rpb1 的 C 端（c）附加了抗原决定簇（FLAG）（标在顶部）的 RNA 聚合酶组装了前起始复合物，然后激发羟基自由基，切割距离 TFIIB 上半胱氨酸 15Å 以内的蛋白质。随后对蛋白质进行 SDS-PAGE，包括前起始复合物蛋白质及蛋白质片段、不替换半胱氨酸的野生型（wt）或无 Fe-BABE（－）的复合物蛋白质及蛋白质片段。蛋白质转膜，用抗 FLAG 抗体（a～c）或 Rpb1 N 端 200 个残基的天然表位制备的抗体分别检测蛋白质印迹。新条带（括号中）表示由羟基自由基切割产生的多肽片段，在对照泳道（“WT”和“－”）未出现新条带。与蛋白质标准（M）比较获得不同片段的长度再结合片段含有 1 个 Rpb1 或 Rpb2 末端的信息可以确定任一侧 4 个氨基酸内的切割位点。在每个括号的旁边标出了切割位点的位置：钳、叉和突起（F/P），活性位点和码头区（A/D）。（e）和（f）在已知酵母 RNA 聚合酶 II 的晶体结构上绘出切割位点［其中 TFIIB 的指域（e）和接头区（f）含有置换的半胱氨酸］。深蓝色表示强切割，浅蓝色表示弱或中度切割。考虑到方法中的误差，以明确的切割位点为中心向外扩展 9 个带颜色的氨基酸。（*Source*：（*a*～*f*）Reprinted from *Cell*，Vol. 119，Hung-Ta Chen and Steven Hahn，“Mapping the Location of TFIIB within the RNA Polymerase II Transcription Preinitation Complex：A Model for the Structure of the PIC，” p. 169-180，fig 2，p. 172. Copyright 2004 with permission from Elsevier.）

小结 对 TFIIB-聚合酶 II 复合物的结构研究表明：TFIIB 通过 C 端域在 TATA 盒处结合 TBP，同时通过 N 端域结合聚合酶 II，这种桥梁作用使聚合酶活性中心粗定位于 TATA 盒下游 25～30bp 处。在哺乳动物中，TFIIB N 端域的环状基序发挥转录起始的精确定位，通过与很靠近活性中心的模板 DNA 单链间的作用而实现。生化研究证明：在前起始复合物中，TFIIB 的 N 端域（具体在指域和接头区）位于靠近 RNA 聚合酶 II 活性中心及 TFIIF 最大亚基的位置。

TFIIH 的结构和功能

TFIIH 是最后一个结合到前起始复合物上的转录因子。在转录起始过程中，可能有两个主要功能：一是使 RNA 聚合酶 II 的 CTD 磷酸化；另一个是使转录起始位点的 DNA 解旋形成“转录泡”。

RNA 聚合酶 II 的羧基末端域 CCTD 的磷酸化

如在第 10 章中所述，RNA 聚合酶以两种生理形式存在：IIA（未磷酸化的）和 IIO（在 CTD 中有大量磷酸化的氨基酸）。未磷酸化的聚合酶 IIA，参与前起始复合物的形成；磷酸化的聚合酶 IIO 催化 RNA 链的延伸。这一现象表明聚合酶磷酸化是在结合前起始复合物与启动子清除两个时段之间。也就是说，聚合酶的磷酸化可能引起聚合酶从起始模式向延伸模式的转化。支持这一推测的证据是聚合酶 IIA 中未磷酸化的 CTD 与 TBP 的结合比聚合酶 IIO 中磷酸化的 CTD 更紧密。因此，CTD 的磷酸化可以破坏聚合酶与 TBP 在启动子区的结合，允许转录进入延伸。但是这一假说也受到一定的质疑，有时体外转录无需 CTD 磷酸化也可以发生。

暂且不提 CTD 磷酸化的重要性，Reinberg 及同事发现 TFIIH 可作为催化磷酸化的候选蛋白激酶。首先，他们发现纯化的转录因子自身就可以让聚合酶 II 的 CTD 磷酸化，将聚合酶 IIA 转化为聚合酶 IIO。图 11.20 中的凝胶阻滞实验证明了这一点。泳道1～6显示加入 ATP 对 DAB、DABPolF、DABPolFE 复合物的迁移率都没有影响，但加入 TFIIH 形成 DABPolFEH 复合物后，ATP 可使复合物的迁移率降低。是什么引起了这种变化？一种可能性是复合物中的一个转录因子使聚合酶磷酸化了。的确，Reinberg 及同事从低迁移率的复合物中分离出的聚合酶是磷酸化的聚合酶 IIO，但是最初加入的是聚合酶 IIA。显然，某个转录因子行使了磷酸化功能。

接下来，Reinberg 及同事直接证明 TFIIH 制备物可使聚合酶 IIA 磷酸化。为此，他们将纯化的聚合酶 IIA 和 TFIIH 及 [γ-^{32}P] ATP 在 DNA 结合状态下温育。如图 11.21a 显示有少量聚合酶发生了磷

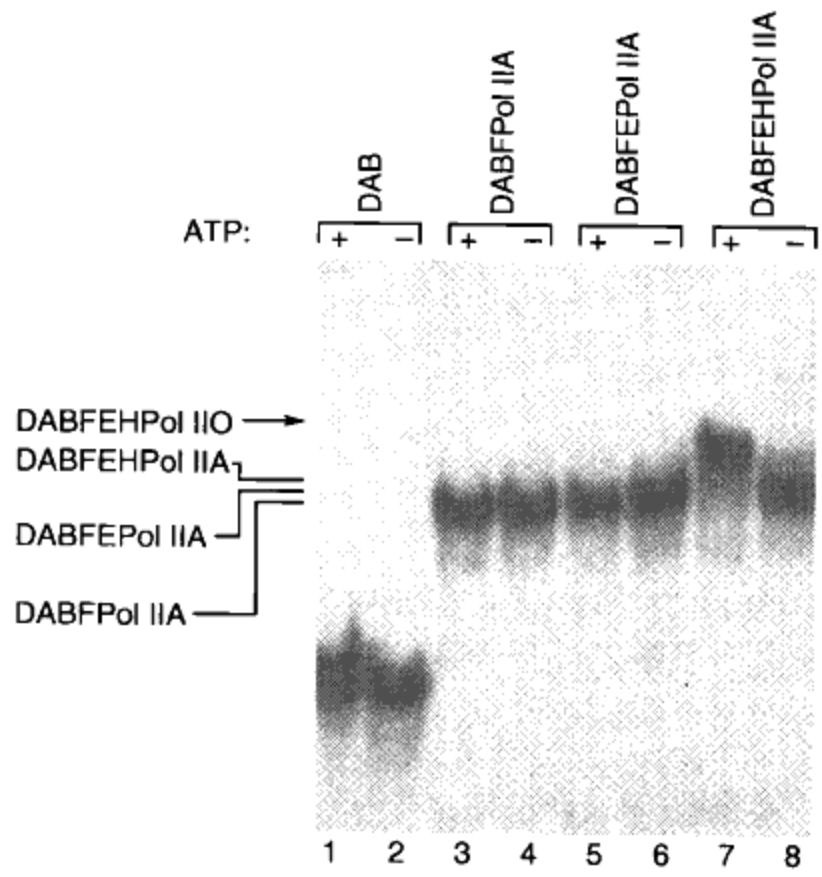


图 11.20 前起始复合物的磷酸化。Reinberg 及同事在有或无 ATP 条件下，将不同的前起始复合物从 DAB 到 DABPolFEH（如顶部所示）进行凝胶阻滞实验。仅当 TFIIH 存在时，ATP 才能改变复合物的迁移率（泳道 7 和 8）。最简单的解释是 TFIIH 促进了聚合酶 IIA 磷酸化为聚合酶 IIO。（*Source*：Lu，H.，I. Zawel，L. Fisher，J. M. Egly，and D. Reinberg，Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II. *Nature* 358 (20 Aug 1992) p. 642，f. 1. Copyright © Macmillan Magazines Ltd.）

酸化。因此，TFIIH 本身可使聚合酶 II 磷酸化，而其他所有因子在一起也不能引起磷酸化。然而，这些因子共同作用可在很大程度上增强 TFIIH 磷酸化的能力。泳道 6～9 显示在逐渐加入其他因子时 TFIIH 的作用结果。每增加一种因子，聚合酶的磷酸化效率和聚合酶 IIO 的积累都有所增加。由于聚合酶 IIO 的最大标记量是在加入 TFIIE 时，因此研究人员又对在 TFIIE 有或无情况下 TFIIH 的磷酸化随时间的变化进行了研究。图 11.21b 显示，当 TFIIE 存在时，IIa 亚基转换为 IIo 亚基的效率高得多。图 11.21c 以曲线图展示了相同结果。

我们知道聚合酶 IIa 亚基的 CTD 是磷酸化的位点，因为无 CTD 的聚合酶 IIB 不能被 TFIIDBFEH 复合物磷酸化，而聚合酶 IIA 以及聚合酶 IIO（在较低程度上）都可以被磷酸化（图 11.22a）。还有，磷酸化产生一个与 IIO 共电泳的多肽，而 IIO 是有 CTD 的。为了直接证明 CTD 的磷酸化，Reinberg 及同事用胰凝乳蛋白酶（chymotrypsin）切割磷酸

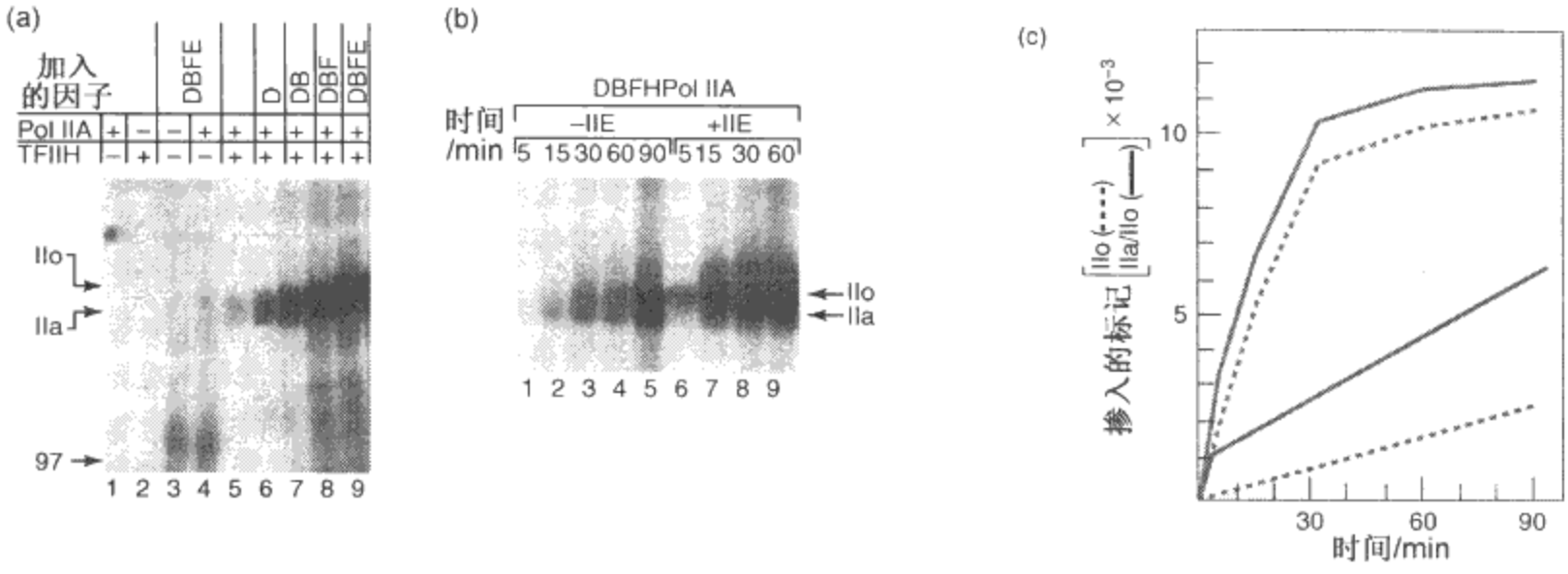


图 11.21　TFIIH 使 DNA 聚合酶 II 磷酸化。（a）Reinberg 及同事将聚合酶 IIA 与各种转录因子的混合物进行温育，如图上部所示。在所有反应中加入［γ-^{32}P］标记的 ATP 促使聚合酶的磷酸化，然后电泳并放射自显影检测。泳道 4，TFIID、TFIIB、TFIIF 和 TFIIE 并不足以引起磷酸化。泳道 5～9 证明仅 TFIIH 就能够引起部分聚合酶的磷酸化，但其他因子可增强 TFIIH 的磷酸化。TFIIE 的加入强烈促进聚合酶 IIa 亚基转化为 IIo 亚基。（b）聚合酶磷酸化的时间进程曲线。在 TFIIE 有或无情况下，利用 TFIID、TFIIB、TFIIF 和 TFIIH 对聚合酶进行相同的磷酸化实验，如顶部所示。实验分别进行了 60min 和 90min，在不同时间取样，如顶部所示。右边箭头指示两种形式的亚基位置。注意 TFIIE 存在时聚合酶磷酸化很快。（c）图是（b）图数据的曲线图示。绿色和红色曲线分别表示 TFIIE 有或无条件下的磷酸化情况。实线对应磷酸化的聚合酶亚基 IIa 和 IIo，虚线对应 IIo。（*Source*：Adapted from Lu，H.，I. Zawel，L. Fisher，J.-M. Egly，and D. Reinberg，Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II. *Nature* 358（20 Aug 1992）p. 642，f. 2. Copyright © Macmillan Magazines Ltd.）

化的聚合酶，该酶可切掉 CTD。电泳酶切产物的放射自显影结果显示有一个标记的 CTD 片段（图 11.22b），表明标记的磷酸已经掺入聚合酶 II 最大亚基的 CTD 中，而亚基的其他部分未被标记。

为了证明没有 RNA 聚合酶 II 的亚基在激酶反应中起作用，Reinberg 及同事克隆了一个嵌合基因，表达包括 CTD、转录因子 GAL4 的 DNA 结合域和谷胱甘肽-S-转移酶的融合蛋白。结果显示 TFIIH 自身就可使此融合蛋白的 CTD 区域磷酸化。因此，即使是在缺少其他聚合酶 II 亚基的情况下，TFIIH 的制备物也完全具有激酶活性。

到目前为止，所描述的实验都是在聚合酶（或聚合酶的结构域）结合 DNA 的条件下进行的。这一点是否重要？为了回答这个问题，Reinberg 小组用聚合酶 II 与不同 DNA 进行了激酶分析实验。所用 DNA 分别为：有完整启动子的、仅有 TATA 盒、仅有起始子区、完全没有启动子。结果表明当 TATA 盒或起始子存在时，TFIIH 对磷酸化有很强的促进作用，但在人工合成的 DNA［poly（dI-dC），不含 TATA 盒或起始子］存在时，磷酸化作用非常弱。因此，只有当聚合酶 II 与 DNA 结合时，TFIIH 才能使聚合酶磷酸化。现在知道 TFIIH 的激酶活性是由其两个亚基提供的。

延伸过程中 RNA 聚合酶 II 中 CTD 的磷酸化和去磷酸化　TFIIH 的蛋白激酶在 RNA 聚合酶 II 的 CTD 的 Ser2 和 Ser5 添加磷酸基团，该磷酸化显然是转录起始所必需的。但是 Ser5 磷酸化在延伸的聚合酶中没有发现，显然是在起始向延伸的转换过程中磷酸被移去了。FCP1 是酵母中唯一已知的 CTD 磷酸酶，但对其基因 *fcp1* 突变后，并不能阻止在延伸过程中 RNA 聚合酶 II 上 CTD 的 Ser5 位去磷酸化。因此，一定有其他磷酸酶移走了这些磷酸基团。

有时在延伸过程中，CTD 的 Ser2 的磷酸化也会被去除，引起聚合酶暂停。要使延伸继续，CTD 的 Ser2 必须重新磷酸化。TFIIH 并不参与这一反应，而是酵母中的 P-TEFb 蛋白对其进行再磷酸化。

小结　前起始复合物形成过程中 RNA 聚合酶 II 以低磷酸化的 IIA 形式存在，随后 TFIIH 使 RNA 聚合酶 II 最大亚基的 C 端域（CTD）上的七氨基酸重复序列上的 2 位和 5 位丝氨酸磷酸化，由此产生聚合酶的磷酸化形式（IIO）。在体外 TFIIE 极大地促进这一过程。磷酸化是转录起始所必需的。在由起始向延伸转变过程中，七氨基酸重复序列 5 位丝氨酸失去磷酸化。如果 2 位丝氨酸也失去磷酸化，聚合酶将暂停等待一个非 TFIIH 激酶进行磷酸化。

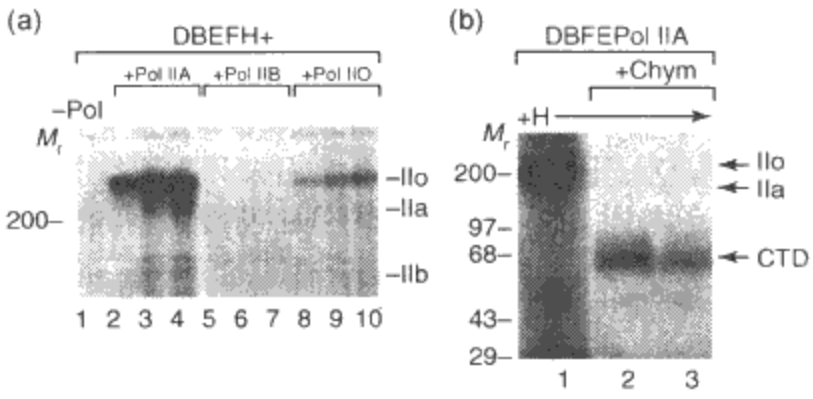


图 11.22　TFIIH 使聚合酶 II 的 CTD 磷酸化。（a）Reinberg 小组在 TFIID、TFIIB、TFIIF、TFIIE 和 TFIIH 及放射性 ATP 存在情况下，逐渐增加聚合酶 IIA、IIB 或 IIO 的量，进行磷酸化实验，如图顶部所示。聚合酶 IIB 缺乏 CTD，因此不能被磷酸化。未磷酸化的聚合酶 IIA 比 IIO 磷酸化程度高，正如预期的。（b）磷酸化 CTD 的纯化。用胰凝乳蛋白酶（Chym）从磷酸化的聚合酶 IIa 亚基中切下 CTD，电泳产物并进行自显影。泳道 1，酶切之前的产物。泳道 2 和 3，酶切之后的产物。CTD 的位置是在另一个实验中确定的。（*Source*：Lu，H.，L. Zawel，L. Fisher，J.-M. Egly，and D. Reinberg. Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II. *Nature* 358（20 Aug 1992）p. 642，f. 3. Copyright © Macmillan Magazines Ltd.）

转录泡的形成　TFIIH 无论在结构或功能上都属于复合蛋白。它含有 9 个亚基，可分解成两个复合物，其中 4 个亚基组成蛋白激酶复合物，另 5 个亚基组成核心 TFIIH 复合物，具有两种独立的 DNA 解旋酶和 ATP 酶（helicase/ATPase）活性。其中一个酶活性在 TFIIH 复合物的最大亚基中，对酵母的生存很重要，当其基因（*RAD25*）突变时，酵母不能存活。Satya Prakash 及同事证明这种解旋酶活性对于转录非常重要。首先，他们在酵母细胞中过量表达了 RAD25 蛋白，纯化均一后表明此产物具有解旋酶活性。他们用部分双链 DNA 作为底物，该底物由一条 ^{32}P 标记合成的 41bp DNA 链与单链 M13 DNA 杂交形成（图 11.23a）。在有或无 ATP 情况下，将 RAD25 和底物混合后进行电泳。解旋酶可以从较长的互补链上释放出具有较高电泳迁移率的标记 DNA 片段，在凝胶下部显示。如图 11.23b 所示，RAD25 具有不依赖 ATP 的解旋酶活性。

接下来，Prakash 及同事发现在含有温度敏感 *RAD25* 基因（*rad25*-ts_{24}）的细胞中，转录也呈温度敏感型。图 11.24 显示以少 G 盒 DNA（第 5 章）为模板的体外转录结果。此模板在一个 400bp 的区域（对应的非模板链上不含 G）上游含有酵母的 TATA 元件。在 ATP、CTP 和 UTP（无 GTP）存在时的转录起始（或终止）显然是在少 G 区域内的两个位点间进行的，并由此产生两个分别为 375nt 和 350nt 的转录物。转录必然在少 G 盒的末端终止，因为超过此位点的链延伸需要 G，但无 G 可用（较短的转录物可能是少 G 盒区域内转录提前终止产生的，而非来自不同起始位点的转录物）。图 11.24a 显示在许可温度（22℃）下转录 0～10min 的结果。很明显，即使在低温下，*rad25*-ts_{24} 突变体细胞提取物也比野生型（*RAD25*）提取物的转录能力弱。图 11.24b 显示在非许可温度（37℃）下的转录结果。温度的升高使 *rad25*-ts_{24} 突变体提取物的转录完全失活。因此，*RAD25* 产物（TFIIH 的 DNA 解旋酶）对转录是必需的。

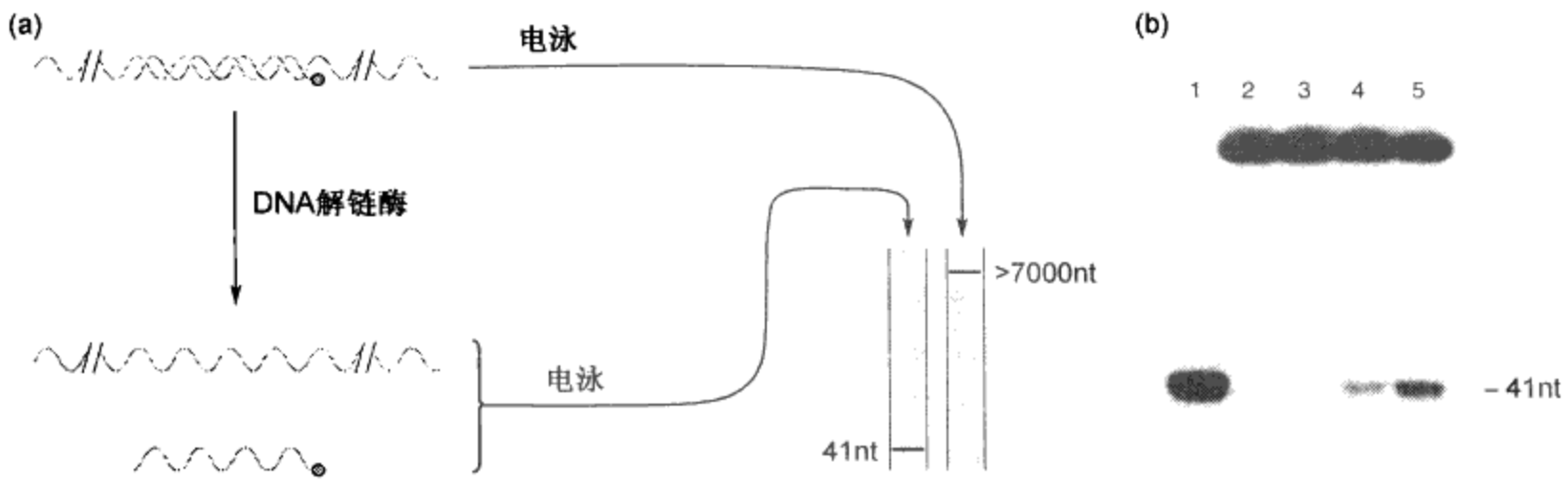


图 11.23　TFIIH 的解旋酶活性。（a）解旋酶分析。标记的 41nt DNA 片段（红色）与较长的未标记单链 M13 噬菌体 DNA（蓝色）杂交，DNA 解旋酶可解开这一短螺旋，从较大的互补链中释放标记的 41nt DNA。通过电泳可区别短 DNA 与其互补链。（b）解旋酶分析结果。泳道 1，热变性底物；泳道 2，未加蛋白；泳道 3，20ng 的不含 ATP 的 RAD25；泳道 4，10ng 的加有 ATP 的 RAD25；泳道 5，20ng 加有 ATP 的 RAD25。（*Source*：（*b*）Gudzer，S. N.，P. Sung，V. Bailly，L. Prakash，and S. Prakash，RAD25 is a DNA helicase required for DNA repair and RNA polymerase II transcription. *Nature* 369（16 June 1994）p. 579，f. 2c. Copyright © Macmillan Magazines Ltd.）

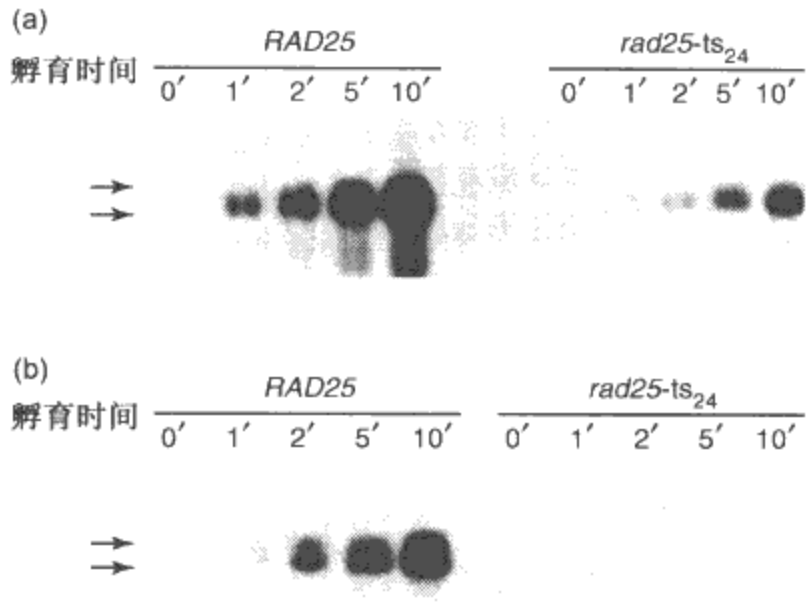


图 11.24　TFIIH 的 DNA 解旋酶基因产物（RAD25）是酵母转录所必需的。Prakash 及同事分别在（a）许可和（b）非许可温度下检测野生型（*RAD25*）和温度敏感突变型（*rad25-ts*$_{24}$）细胞提取物对少 G 盒模板的转录。转录在有 ATP、CTP 和 UTP（无 GTP）且其中一种核苷酸被^{32}P 标记情况下进行 0 ～10min。电泳标记产物，放射自显影检测条带。提取物（*RAD25* 或 *rad25-ts*$_{24}$细胞）的来源，及温育时间在顶部标出。左边的箭头指示两种少 G 盒转录产物的位置。可以看到当 TFIIH 的 DNA 解旋酶（RAD25）为温度敏感型时，转录即呈温度敏感型。（*Source*：Gudzer, S. N., P. Sung, V. Bailly, L. Prakash, and S. Prakash, RAD25 is a DNA helicase required for DNA repair and RNA polymerase II transcription. *Nature* 369（16 June 1994）p. 580, f. 3 b-c. Copyright © Macmillan Magazines Ltd.）

转录的哪一步需要 DNA 解旋酶活性呢？一连串实验结果引导研究者从以下思路来探讨答案。不像Ⅰ类和Ⅲ类基因，Ⅱ类基因的转录需要 ATP（或 dATP）的水解。当然，所有 4 种核苷酸包括 ATP 的 α-β-磷酸酯键，在所有转录过程中均被水解，但Ⅱ类转录需水解 ATP 的 β-γ-磷酸酯键。问题是哪一步需要水解 ATP？很自然地联想到 TFIIH，因为它的两种酶活性（CTD 激酶和 DNA 解旋酶）都需要 ATP 水解。答案是 TFIIH 的解旋酶活性需要 ATP。有力的证据是在 CTD 磷酸化过程中，GTP 可以替代 ATP，而在转录过程中 GTP 则不能满足 ATP 水解的要求。因此，除 CTD 的磷酸化外，转录的某个过程还需要 ATP 的水解，而最可能的选择就是 DNA 解旋酶。

现在再回到主题，转录的哪一步需要 DNA 解旋酶活性？最可能的答案就是启动子清除。在第 6 章定义转录起始时包括了启动子清除，但启动子清除也可看做一个独立事件，它是转录起始和延伸的分界线。James Goodrich 和 Tjian 提出了这样的问题：TFIIE 和 TFIIH 是起始所必需还是启动子清除所必需？为了回答这个问题，他们设计了一个实验可测定中断转录物（三核苷酸）。中断转录物的出现表明，有功能的转录起始复合体已经形成，包括 DNA 局部解链和第一个磷酸二酯键的形成。他们发现 TFIIE 和 TFIIH 并不是中断转录物所必需的，而 TBP、TFIIB、TFIIF 和 RNA 聚合酶Ⅱ却是必需的。因此，TFIIE 和 TFIIH 不是转录起始所必需的，至少在启动子清除之前不需要，但是启动子 DNA 完全解链时需要 TFIIH。如果人 TFIIH 的最大亚基突变，就丧失了其中的 DNA 解旋酶功能，启动子 DNA 就不能完全打开，这又阻止了启动子清除，后面将会予以解释。

这些发现表明了 TFIIE 和 TFIIH 对启动子清除或 RNA 延伸单独或同时都需要的可能性。为了区分这些可能性，Goodrich 和 Tjian 研究了 TFIIE 和 TFIIH 对延伸过程的影响。通过剔除在第 17 位但不是之前的核苷酸，在超螺旋模板上起始（没有 TFIIE 和 TFIIH）了转录并进行到第 16 位核苷酸（用超螺旋模板是因为在这种模板上体外转录不需要 TFIIE 和 TFIIH，也不需要 ATP）。然后用限制酶酶切使模板线性化，再加入 ATP，在有或无 TFIIE 和 TFIIH 情况下继续转录。结果发现 TFIIE 和 TFIIH 的有无对延伸无影响。由于 TFIIE 和 TFIIH 对起始和延伸都无影响，Goodrich 和 Tjian 认为 TFIIE 和 TFIIH 在启动子清除这一步是必需的。图 11.25 总结了以上情况，下一个段落将予以讨论更新近的数据。

Tjian 等推测 TFIIH 的 DNA 解旋酶活性直接作用于起始子位点的 DNA 并使其解链。但 Tae-Kyung Kim 等在 2000 年进行的交联实验表明，TFIIH（具体为含有启动子解链活性的 DNA 解旋酶亚基）与 +3 至 +25 之间或更下游位点的 DNA 产生交联，TFIIH 作用位点位于第一个转录泡（−9 至 +2 位）的下游。另外，TFIIE 与转录泡区域发生交联，TFIIB、TFIID 和 TFIIF 与转录泡的上游区域交联，而 RNA 聚合酶的交联区则包含了所有其他转录因子的全部区域。这一发现揭示，TFIIH 中的 DNA 解旋酶并不与第一个转录泡接触，因此不可能通过对 DNA 的直接解旋来产生转录泡。加入 ATP 对转录泡上游区的作用没有影响，但确实能够干扰转录泡内及其下游的作用。

从以前的研究知道，TFIIH 解旋酶负责形成转录泡，但此处的交联实验则表明它不直接使转录泡的 DNA 解旋。那么它如何产生转录泡呢？Kim 及同事认为解旋酶可以像分子“扳手”那样使下游 DNA

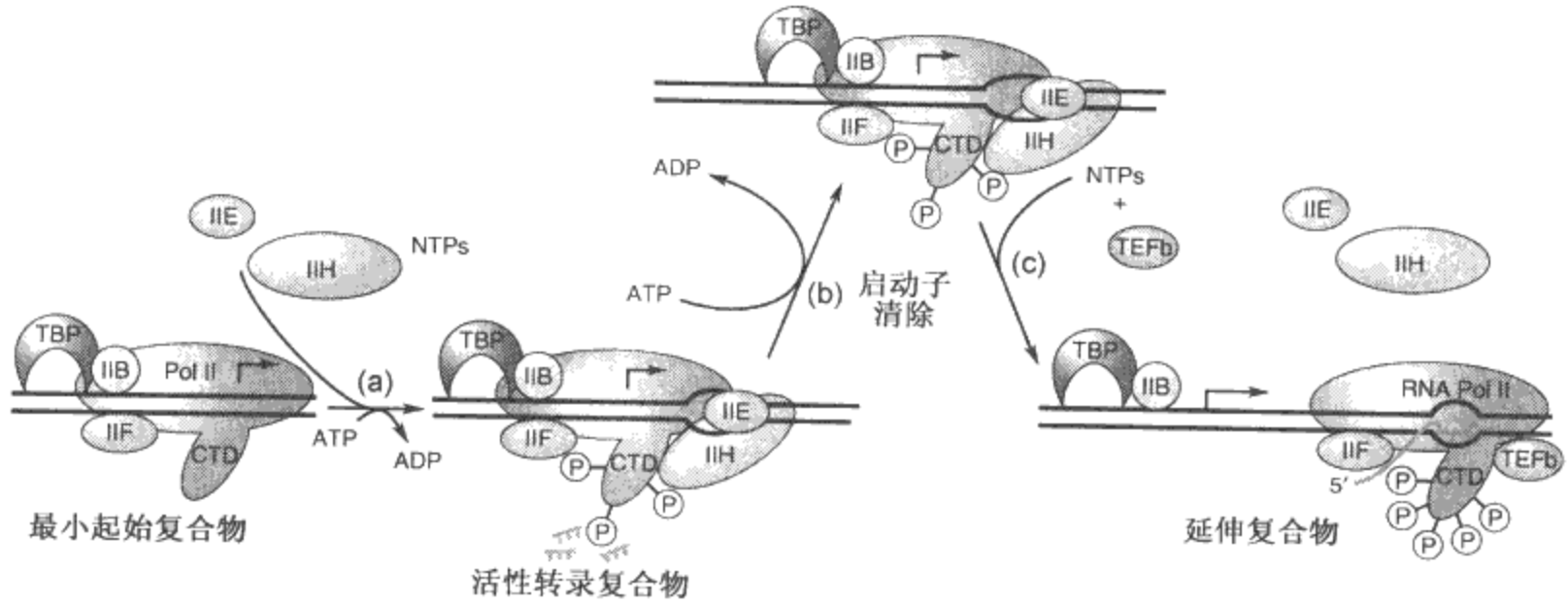


图 11.25　通用转录因子参与转录起始、启动子清除、延伸过程的模型。（a）TBP（或 TFIID）与 TFIIB、TFIIF 和 RNA 聚合酶 II 共同在起始子处形成最小起始复合物。添加 TFIIH、TFIIE、ATP 使 DNA 在起始区解链，RNA 聚合酶最大亚基的 CTD 产生部分磷酸化。这些过程导致中断转录物的产生（洋红色），聚合酶随后在+10～+12 位置停止转录。（b）随着 ATP 提供能量，TFIIH 的 DNA 解旋酶引起 DNA 进一步解旋，转录泡增大。这一过程释放聚合酶并使其跨过启动子区。（c）随着聚合酶 CTD 在 TEFb 作用下进一步磷酸化及 NTP 的不断添加，延伸复合物继续使 RNA 链延长，TBP 和 TFIIB 仍然留在启动子上。在延伸中不需要 TFIIE 和 TFIIH，它们从延伸复合物上解离下来。（*Source*：Adapted from Goodrich, J. A. and T. Tjian, 1994. Transcription factors IIE and IIH and ATP hydrolysis direct promoter clearance by RNA polymerase II. *Cell* 77：145-156.）

解旋。因为 TFIID 和 TFIIB（或其他蛋白质）紧紧地结合在转录泡上游的 DNA 上，而且加入 ATP 后仍不放松，转录泡下游 DNA 的解旋将在上下游之间产生一种张力，从而打开转录泡的 DNA。这使聚合酶起始转录并向下游移动 10～12bp。但以前的研究表明：聚合酶将在此处停顿，等待 TFIIH 的进一步帮助，使下游 DNA 进一步扭曲以加长转录泡，最终释放停止的聚合酶从而使启动子得以清除。

图 11.25 是示意图，以便理解 TFIIH 在 CTD 磷酸化和 DNA 解旋过程中的作用。真正的前起始复合物结构要复杂得多。Kornberg 及同事根据以前对 TFIIE-聚合酶 II、TFIIF-聚合酶 II 以及 TFIIE-TFIIH 复合物的结构研究，模拟了所有通用转录因子（TFIIA 除外）在前起始复合物中的位置（图 11.26）。TFIIF 第二大亚基（Tfg2）与细菌 σ 因子同源，而且在启动子上与 σ 因子几乎处于一致的位置。图中 Tfg2 上与大肠杆菌 σ 因子的 2、3 区同源的两个域已标记为“2”和“3”。TFIIE 位于聚合酶活性中心下游 25bp 处，以履行召集 TFIIH 的职责。而 TFIIH 则通过直接或间接诱导反向超螺旋行使其作为分子扳手的 DNA 解旋酶功能，以打开启动子区的 DNA 链。

小结　TFIIE 和 TFIIH 对形成开放启动子复合物或延伸并非必需，但在启动子清除中起作用。TFIIH 的解旋酶活性是转录必需的，推测是由于此活性使启动子区的 DNA 完全解链从而促进启动子清除。

中介物复合物和 RNA 聚合酶 II 全酶

另一类被称为中介物（mediator）的蛋白质集合也可以认为是通用转录因子，因为它是大多数 II 类前起始复合物的组成部分。不像其他的通用转录因子，中介物本身并不参与转录起始，在第 12 章将看到，它在被激活的转录过程中起作用。在酵母中首先发现了含有 20 条多肽的中介物，随后在人类也发现了超过 20 条多肽的中介物大复合物，其中只有少部分多肽与酵母中介物具有明显同源性。

到目前为止我们讨论 II 类启动子上前起始复合物的装配都是按一次一个蛋白质进行的，这种情况可能确实发生了。但是有证据表明 II 类转录前起始复合物可以通过将预先形成的 **RNA 聚合酶 II 全酶**（RNA polymerase II holoenzyme）结合到启动子上而装配。全酶包括 RNA 聚合酶、一套通用转录因子和中介蛋白质。

1994 年，Roger Kornberg 和 Richard Young 的实验室的研究提供了有关全酶概念的证据。两个小组分别从酵母细胞中分离到一种复合蛋白，含有 RNA 聚合酶 II 和其他多种蛋白质。Kornberg 及同事用抗全酶某个组分的抗体直接对整个复合物进行

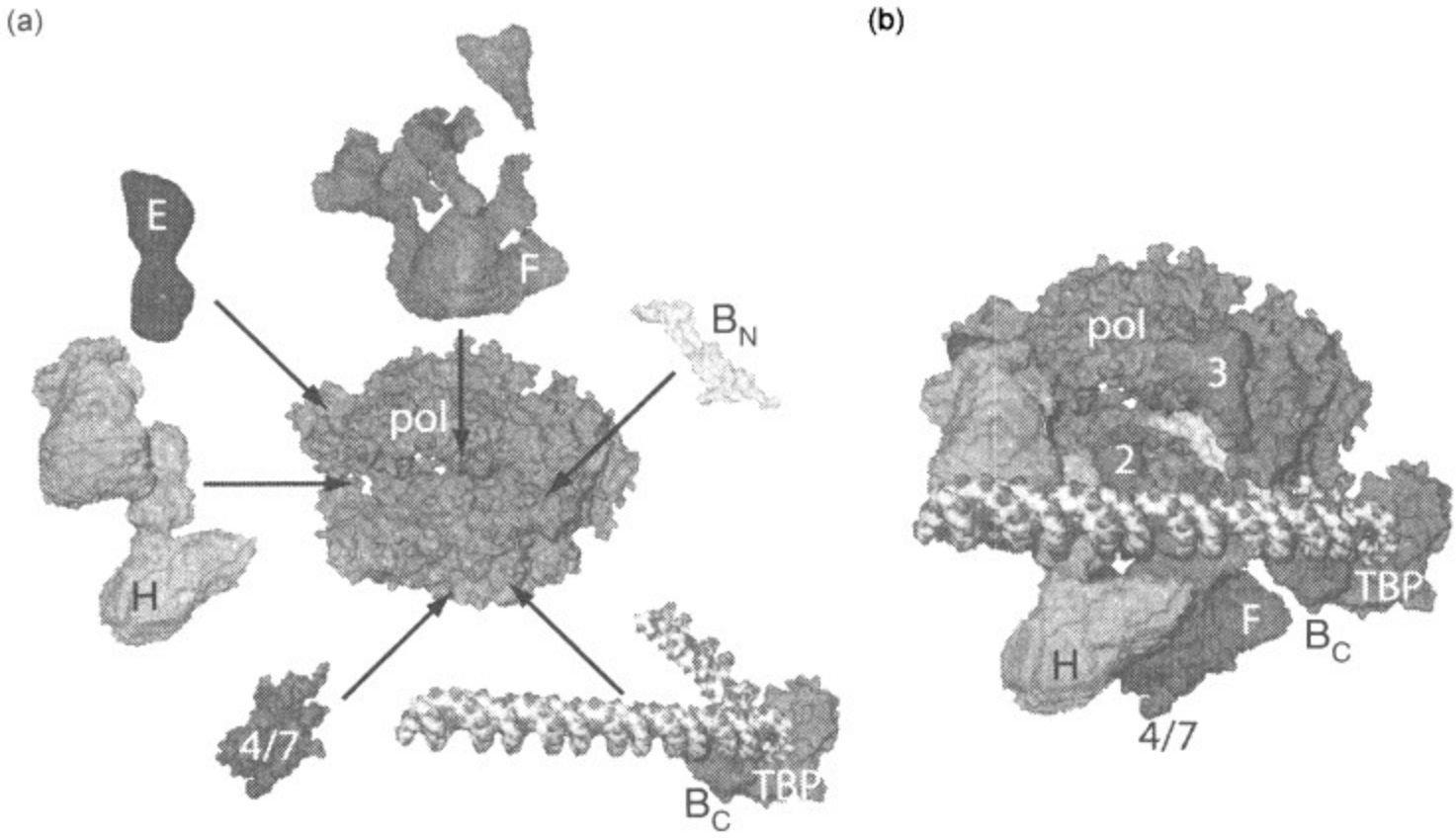


图 11.26　II 类前起始复合物模型。Kornberg 及同事将以前关于启动子 DNA、TFIIF、TFIIE 和 TFIIH 位置的信息加入到他们的 TFIIB-聚合酶 II 复合物晶体结构中，产生了该组合模型。(a) 复合物单个组分的局部分解图。红色（4/7）表示 Rpb4 和 Rpb7，pol（灰色）表示 RNA 聚合酶的其余部分，B_N 和 B_C 分别表示 TFIIB 的 N 端和 C 端域。启动子 DNA 由红白蓝模型表示，由于 TBP 的结合导致了 DNA 的一个大角度弯曲。(b) 完整结构。注意此时转录泡还未形成，转录的方向从右至左。(*Source*：(*a-b*) Reprinted with permission from *Science*, Vol. 303, David A, Bushnell, Kenneth D. Westover, Ralph E. Davis, Roger D. Kornberg, "Structural Basis of Transcription: An RNA Polymerase II-TFIIB Cocrystal at 4.5 Angstroms" Fig. 6, p. 986, Copyright 2004, AAAS.)

免疫共沉淀，获得 RNA 聚合酶 II 的亚基、TFIIF 的亚基和其他 17 种多肽。通过加入 TBP、TFIIB、TFIIE 和 TFIIH，就可以完全恢复这个全酶的转录活性。在这里并不需要 TFIIF，因为它已经是全酶的一部分了。

Anthony Koleske 和 Young 用一系列纯化步骤从酵母中分离出含有 RNA 聚合酶 II、TFIIB、TFIIF和 TFIIH 的全酶。只需要 TFIIE 和 TBP 就能在体外进行精确转录，因此该全酶比 Kornberg 等分离的全酶含有更多通用转录因子。Koleske 和 Young 在全酶中也鉴定出一些中介物多肽并称其为 SRB 蛋白（SRB2、SRB4、SRB5 和 SRB6）。SRB 蛋白是由 Young 及同事在进行遗传筛选中发现的，其由来是这样：当删除聚合酶 II 上 CTD 的一部分后，导致 GAL4 蛋白（将在第 12 章详细研究的一个转录激活因子）对转录的无效激活。接着他们开始筛选能够抑制 GAL4 无效刺激的突变体，鉴定出几种抑制子突变型基因，命名为 *SRB*（suppressor of RNA polymerase B），在第 12 章将讨论该抑制的可能机制。事实已足以证明 SRB 蛋白对体内最佳转录激活是必需的，至少在酵母中是这样，而且这些蛋白质也是酵母聚合酶 II 全酶中的中介物复合物的一部分。哺乳动物包括人的全酶已经分离出来。

小结　酵母和哺乳动物细胞具有 RNA 聚合酶 II 全酶，它除了具有聚合酶的亚基外还包括许多多肽。额外的多肽包括一系列通用转录因子（TBP 除外）和中介物。

延伸因子 TFIIS

真核生物转录调控主要发生在转录起始步骤，但至少在 II 类基因也对延伸过程实施一些调控。1987 年，Reinberg 和 Roeder 发现了一个 HeLa 细胞因子，并命名为 **TFIIS**。TFIIS 能显著促进体外转录的延伸。该因子和 IIS 同源，IIS 最初是由 Natori 及同事在埃利希腹水（Ehrlich ascites）肿瘤细胞中发现的。

TFIIS 促进延伸　Reinberg 和 Roeder 在前起始复合物上对 TFIIS 进行了检测，证明它作用于延伸而不是起始过程（图 11.27）。他们把聚合酶 II、DNA 模板和核苷酸一起温育进行转录的起始，然后加入肝素结合自由的聚合酶，以阻断新的起始，最后加入 TFIIS 或缓冲液来测定标记 GMP 掺入 RNA 的速率。图 11.27 显示 TFIIS 显著增强了 RNA 合

成，垂直虚线显示在 6min 时 TFIIS 使 GMP 掺入量提高 2 倍，10min 时达 2.6 倍。很明显，延伸速率的增加更迅速，至少增加了 10 倍。

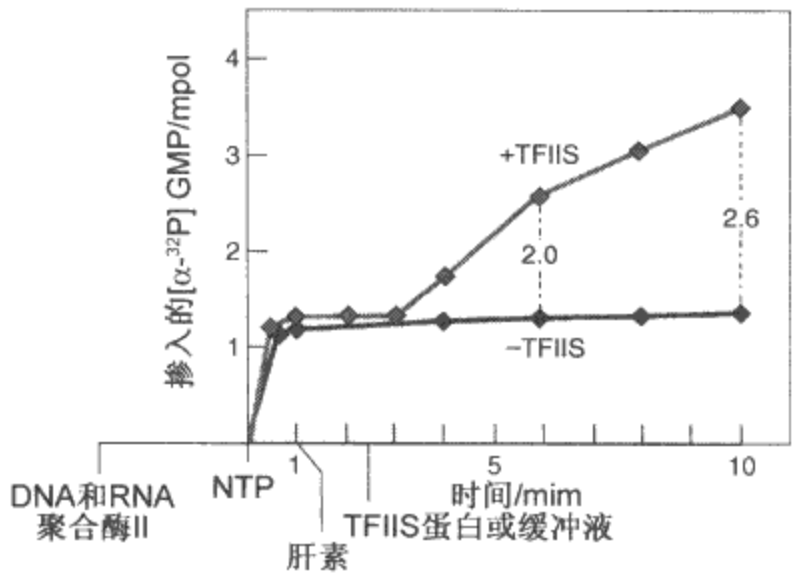


图 11.27　TFIIS 对转录延伸的影响。Reinberg 和 Roeder 构成的延伸复合物概括于图下的时间轴上。在 −3min 时，加入 DNA 和 RNA 聚合酶，在 0 时加入 4 种 NTP 以起始反应，其中 GTP 用 ^{32}P 标记。在 +1min 时，加入肝素结合游离的 RNA 聚合酶，由此开始所有转录复合物应该是延伸复合物了。最后，在 2.5min 时，加入 TFIIS（红色）或作为阴性对照的缓冲液（蓝色）。在标记 GMP 掺入的不同时间对反应混合物取样，测定掺入 RNA 的 GMP 量。垂直虚线显示由 TFIIS 刺激后总 RNA 合成增加的倍数。（*Source*：Adapted from D. Reinberg and R. G. Roeder, Factors involved in specific transcription by mammalian RNA polymerase II. Transcription factor IIS stimulates elongation of RNA chains. *Journal of Biological Chemistry* 262：3333，1987.）

在以上实验中 TFIIS 也存在刺激转录起始的可能性。为了研究这一点，Reinberg 和 Roeder 重复了上述实验，但在加入肝素之前，即零时开始温育时就加入 TFIIS，如果 TFIIS 真能像促进延伸一样对起始产生影响，那么此实验会产生比前次更强的增强作用。但图 11.27 表明：TFIIS 在这两个实验中产生的作用几乎一样。因此，TFIIS 只对延伸产生促进作用。

那么 TFIIS 怎样增强转录的延伸呢？Reinberg 和 Roeder 设计了一个实验，明确显示 TFIIS 通过解除聚合酶的暂停而行使功能。RNA 聚合酶的一个普遍特征是它并不以均一的速度转录，而是会产生暂停现象，有时在重新开始转录前要经历很长时间的暂停。这些暂停倾向于在特定**暂停位点**（pause site）发生，因为这些位点的 DNA 序列使 RNA-DNA 杂交体变得不稳定，并导致聚合酶后退，使新合成 RNA 的 3′端被排入聚合酶的一个孔中，如第 10 章介绍的。如果后退仅限于几个核苷酸，则聚合酶可自行恢复转录。但如果后撤距离过长，聚合酶就需要 TFIIS 因子来帮助其恢复转录。后者是一种被称作**转录停滞**（transcription arrest）的更严重状态，而不是**转录暂停**（transcription pause）。

体外转录中的暂停（或停滞）可以通过对转录物电泳进行检测，可发现比全长转录物短的一系列片段。Reinberg 和 Roeder 发现 TFIIS 能减少短转录片段的量，表明它能最大限度地减少转录的停滞。其他研究人员也证明了这一结论。

Daguang Wang 和 Diane Hawley 在 1993 年证明了 RNA 聚合酶 II 具有内在的弱 RNase 活性，且能被 TFIIS 激活，这一发现及后续研究使研究人员提出了 TFIIS 如何重起转录停滞的假说（图 11.28）。停滞的 RNA 聚合酶后撤太远致使新合成的 RNA 3′端离开酶的活性中心，并通过孔和漏斗排出。在没有 3′端核苷酸加入的情况下，聚合酶停止工作。而 TFIIS 此时激活了 RNA 聚合酶 II 的 RNase 活性，切去新生 RNA 被排出的部分，在酶的活性中心产生了一个新的 3′端。

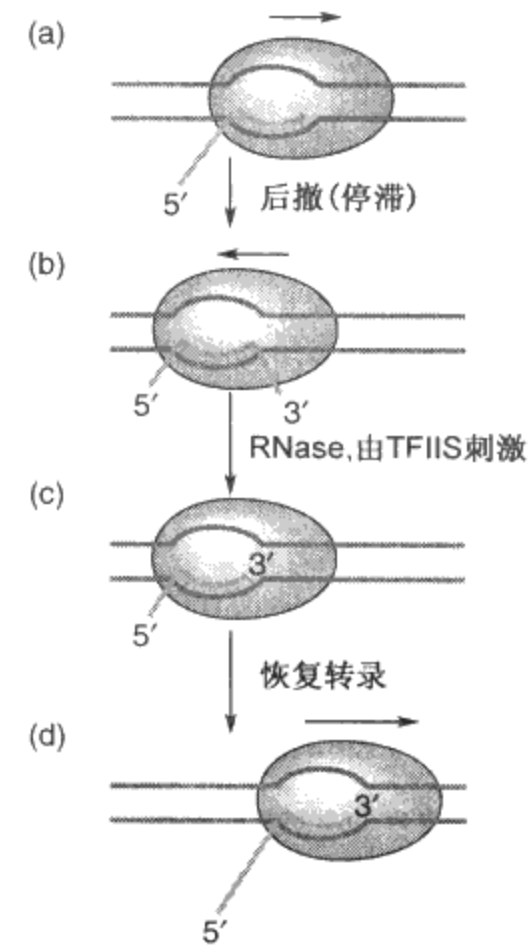


图 11.28　TFIIS 对转录停滞的逆转模型。(a) 从左到右转录 DNA 的 RNA 聚合酶 II 刚刚停下来。(b) 聚合酶向左后撤，露出新生 RNA 的3′端，使之不在酶的活性位点。由此引起转录停滞，RNA 聚合酶自己不能恢复。(c) 聚合酶潜在的核酸酶活性被 TFIIS 激活，切除了新生 RNA 的 3′端。(d) 游离的 RNA 3′端回到活性中心，聚合酶恢复转录。

TFIIS如何将一个通常是合成RNA的酶转变为裂解RNA的酶呢？Patrick Cramer及同事获得的RNA聚合酶II-TFIIS复合物的X射线晶体结构有助于解答这一问题。图11.29显示基于此复合物晶体结构的剖面图。TFIIS由3个域组成，其中之一呈锌带特征，此锌带结构与被挤出的RNA同时位于聚合酶II的孔和漏斗中。在锌带的尖上有两个酸性氨基酸残基与聚合酶活性中心的金属A非常靠近，有利于氨基酸的酸性侧链螯合第二个镁离子，该离子与第一个镁离子共同参与核酸酶活性。

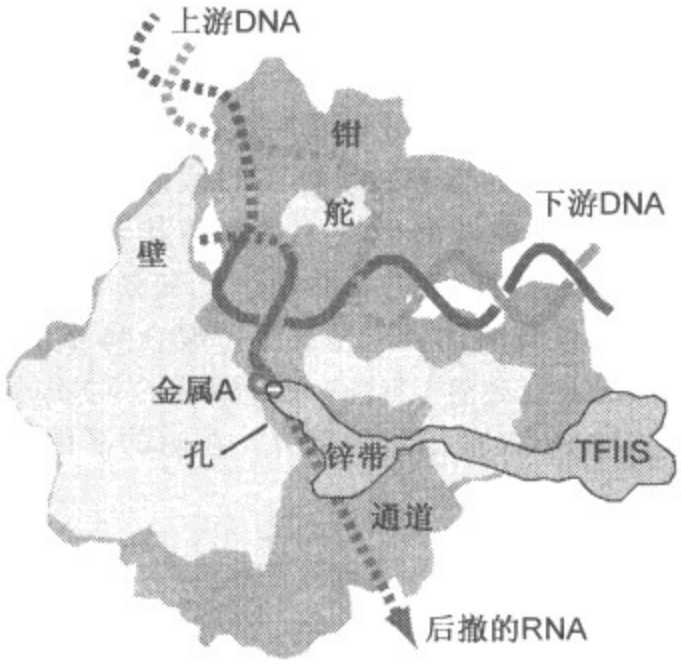


图11.29　被阻滞的酵母RNA聚合酶II-TFIIS复合体的剖面观。聚合酶后退，将新合成的RNA 3′端（红色）排出酶的活性中心进入孔道中。TFIIS（橙色）的锌带结构也位于此孔道中，且其顶部含有两个酸性残基（绿色），靠近聚合酶活性中心的金属A（紫红色），有利于两个酸性氨基酸螯合第二个金属离子，该离子与第一个离子一起配合参与核酸酶活性，切割被挤出的RNA 3′端。（*Source*：Reprinted from *Cell*，Vol. 114，Conaway et al.，"TFIIS and GreB：Two Like-Minded Transcription Elongation Factors with Sticky Fingers，" fig. 1，p. 272-274. Copyright 2003，with permission from Elsevier. Image courtesy of Joan Weliky Conaway and Patrick Cramer.）

这样看来TFIIS不是通过结合在聚合酶表面以改变其内部构型而改变RNA聚合酶的活性，而是直接进入酶的活性中心并积极参与其催化作用。这一推测得到了有力的支持：因为发现了大肠杆菌蛋白**GreB**，它在重新启动转录停滞中具有与TFIIS相同的作用。这两种蛋白质没有同源性，因此不太可能起源于共同的祖先。然而，GreB有一个卷曲螺旋域延伸至大肠杆菌RNA聚合酶的排出RNA的出口通道，这与TFIIS中锌带具有相同的结构和作用方式。而且在GreB卷曲螺旋顶部与聚合酶活性中心金属离子相邻处，有两个酸性氨基酸残基可能与TFIIS中对应部分在核酸酶催化过程中具有相同的功能。这一明显的功能趋同进化（convegent evolution）结果，为所推测的TFIIS功能的有效性提供了证据。

有趣的是，有报道认为一种起始因子（TFIIF）在延伸中也起作用。它不像TFIIS那样解除聚合酶在特定位点的停滞，而是限制聚合酶在DNA任意位点的瞬时停顿。

小结　转录可在延伸水平调控。TFIIS因子通过限制聚合酶在不连续位点的转录停滞而刺激延伸。TFIIF通过插入RNA聚合酶的活性中心，激发其核酸酶活性切割排出的引起停滞的新生RNA的3′端而发挥其功能。TFIIF显然是通过限制瞬时停顿而刺激延伸。

TFIIS促进转录物的校正　TFIIS不仅解除暂停，而且具有校正转录物的功能，其机制可能是激活RNA聚合酶内在的RNase活性，清除错配的核苷酸。根据图11.30a所示的步骤，Diane Hawley及同事测定了TFIIS的校正功能。首先，他们分离了在邻近启动子多个位点发生暂停的未标记延伸复合体。接着，在放射性UTP存在条件下使复合物步移至规定位置（第6章），并使复合物中的RNA被标记。然后加入ATP或GTP使RNA延伸一个碱基至+43位点。在该位置需要掺入的碱基是A，但如果只有G，聚合酶也会低效地将其掺入。实际上他们发现极纯的GTP里仍含有少量ATP，因此即使加入的只是GTP，仍不影响GMP和AMP以相等的数量掺入+43位。接下来分两组：一组用T1 RNase（在G之后切割）切割RNA产物；另一组用4种核苷酸驱赶获得全长标记的RNA，再用T1 RNase切割产物。最后电泳酶切产物，放射自显影观察被标记的产物。

因为电泳可明确区分以A和G结尾的末端7碱基序列，因此他们只需用T1 RNase切割转录物就能测定GMP和AMP掺入+43位的相对量。在图11.30b中，泳道1显示无追踪的结果，G结尾的末端七聚体（错误掺入G的结果）的量大致等于以A结尾末端七聚体和AC结尾末端八聚体之和的量（正确掺入A或C，因GTP底物中有其他核苷酸污染）。泳道2和3分别显示在TFIIS有或无条件下追踪的结果。用T1 RNase切割追踪的全长转录物，产生以Gp结尾的末端七聚体（来自有错误掺入G的全长转录物），或以Gp结尾末端的十聚体（来

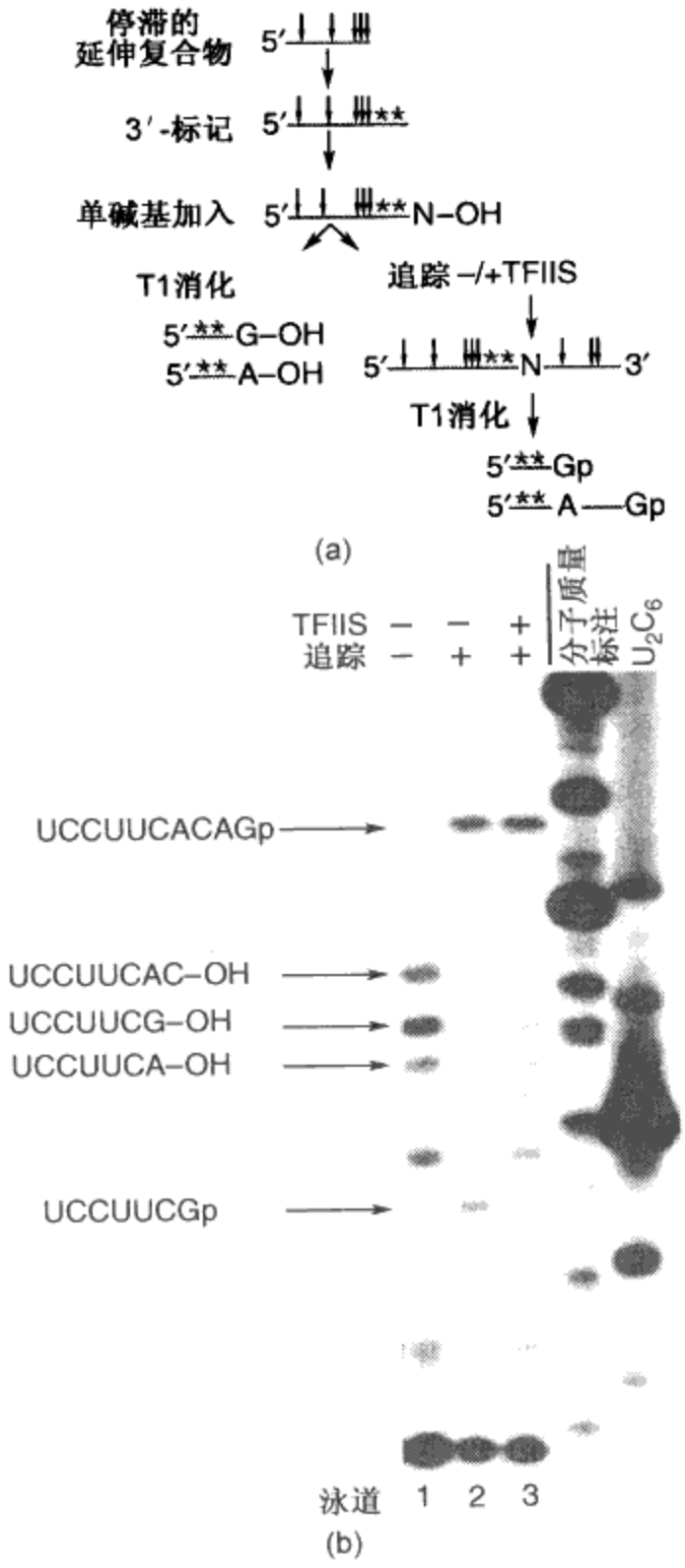


图 11.30　TFIIS 通过 RNA 聚合酶 II 促进校正。（a）实验流程图。Howley 及同事从短的延伸复合体开始，在 α-^{32}P UTP 存在条件下通过步移聚合酶使短的转录物在 3′端标记，随后加入 GTP 迫使 G 错误掺入需要 A 的＋43 位。然后用 T1 RNase 消化标记的转录物以测定错误掺入 G 的量（左边），或加入 4 种核苷酸驱赶转录物获得全长。再用 T1 RNase 切割转录物以测定对＋43 位 G 的校正情况。（b）实验结果。将（a）中的酶切产物进行电泳并放射自显影。泳道 1，未追踪的转录物。左边箭头分别指示错误掺入 G 的 7 聚体（UCCUUCG$_{OH}$）、正确结合 A（或 A 和 C）的 7 聚体（UCCUUCA）和 8 碱基序列（UCCUUCAC）。泳道 2 和 3 显示 TFIIS 无（泳道 2）或有（泳道 3）时驱赶转录的酶切产物。左边箭头分别指示错误掺入 G 的 7 聚体（UCCUUCGp）、在＋43 位正确结合 A 和通过校正将 G 换为 A 的 10 聚体（UCCUUCACAGp）。TFIIS 使所有可检测到的错误掺入 G 都能被去除。（*Source*：（*b*）Thomas，M. J.，A. A. Platas，and D. K. Hawley，Transcriptional fidelity and proofreading by RNA Polymerase II. *Cell* 93 (1998) f. 4，p. 631. Reprinted by permission of Elsevier Science.）

自将 G 校正为 A 的全长转录物）。在无 TFIIS 条件下进行追踪时，RNA 产物中有大量错误掺入的 G（泳道 2 中与指示七聚体 UCCUUCGp 的箭头相对的条带）。但大多数产物都是十聚体（箭头指示的 UCCUUCACAGp），表明即使无 TFIIS 聚合酶也能进行部分校正。当加入 TFIIS 进行追踪时，七聚体消失，所有标记产物均为十聚体。因此，TFIIS 能促进转录物的校正。

当前的校正模型是：聚合酶不仅在掺入错误核苷酸时做出暂停的响应，而且后退并排出新合成的 RNA 3′端从而引起转录停滞。随后由 TFIIS 激活聚合酶潜在的 RNase 活性，切除被排出的 RNA 末端及错配的核苷酸，使聚合酶恢复转录。

小结　TFIIS 促进校正，即错配核苷酸的改正。可能是通过激活 RNA 聚合酶中的 RNase 活性，切除错配的核苷酸并代以正确的核苷酸。

11.2　I 类转录因子

在 rRNA 启动子上的前起始复合物比前面讨论的聚合酶 II 的要简单得多。此复合物中除聚合酶 I 之外，只包含两个转录因子：一个为核心结合因子（core-binding factor），在人类称为 **SL1**，在一些其他生物称为 **TIF-IB**；另一个是 UPE 结合因子，在哺乳动物称为**上游结合因子**（upstream-binding factor，UBF），在酵母称为**上游激活因子**（upstream activating factor，UAF）。SL1（或 TIF-IB）和 RNA 聚合酶 I 是基本转录激活所必需的。实际上，核心结合因子对聚合酶 I 结合启动子是必需且充分的。UBF（或 UAF）是结合 UPE 的**装配因子**（assembly factor），通过使 DNA 剧烈弯曲帮助 SL1 结合到核心启动子上，所以也叫做**结构转录因子**（architectural tran-

scription factor）（第 12 章）。人和非洲爪蟾在转录 I 类基因时绝对依赖 UBF，而其他生物（包括酵母、大鼠、小鼠）在没有组装因子时也可进行部分转录。还有一些物种，如卡氏棘阿米巴变形虫（*Acanthamoeba castellanii*）几乎不依赖于组装因子。

核心结合因子

1985 年，Tjian 及同事在将 HeLa 细胞提取物分成两种功能成分的过程中发现了 SL1 因子：一种成分有 RNA 聚合酶 I 活性，但是不能在体外对人 rRNA 基因精确地起始转录；另外一种成分自身没有聚合酶活性，但可以指导前一组分在人 rRNA 模板上精确地起始转录。而且，转录因子 SL1 表现出物种特异性，即它可以区别人和小鼠的 rRNA 启动子。

以上实验使用的都是不纯的聚合酶 I 和 SL1 因子。利用高纯度成分进行的深入研究显示，人类 SL1 因子不能独立激活人聚合酶 I 结合 I 类启动子并起始转录。随后发现它需要 UBF 帮助。与此不同的是，阿米巴变形虫的 I 类基因在无 UBF 时也可以转录。

由于无 UBF 时用核心结合因子 SL1 进行的人 I 类基因的转录效果很差，因此不宜用人类系统研究核心结合因子募集聚合酶 I 结合启动子的功能。而卡氏棘阿米巴变形虫是较好的选择，它对 UPE-结合蛋白几乎没有依赖性，可以单独研究核心结合因子。Marvin Paule 和 Robert White 研究了该系统，发现核心结合因子（TIF-IB）可以募集聚合酶结合到启动子上并在合适的位置激活转录起始。而聚合酶结合 DNA 的实际序列好像与此无关。

他们在 TIF-IB 结合位点和正常转录起始位点间插入或删除不同数目的碱基，制备了突变体模板。这会让我们想起第 10 章 Benoist 和 Chambon 研究 II 类启动子的实验。在那个实验中，删除位于 TATA 盒与正常转录起始位点间的碱基不改变转录的强度，也不改变相对于 TATA 盒的转录起始位点。在所有情况下，转录均发生在距 TATA 盒下游约 30bp 处。

对于 I 类启动子，Paule 及同事也得出了相似的结论。他们发现在 TIF-IB 结合位点和正常转录起始位点之间增加或删除多至 5 个碱基时转录仍能发生，而且转录起始位点随着增加或删除的碱基数而向上游或下游移动（图 11.31），当增加或删除多于 5 个碱基时转录被阻断。于是 Paule 及同事得出结论：TIF-IB 与聚合酶作用并将其定位于下游若干个碱基处起始转录。

聚合酶结合 DNA 的准确序列不重要，因为在各突变 DNA 中它们不尽相同。为了确定聚合酶在各突变体中结合 DNA 的位点与 TIF-IB 结合位点之间有相等间距，Paule 及同事用野生型和不同突变体为模板进行了 DNase 足迹实验。结果表明不同模板的足迹基本无区别，更加证实了无论 DNA 序列如何，聚合酶都以相同位距结合 DNA 的结论。这与 TIF-IB 结合靶 DNA 并通过直接的蛋白质-蛋白质相互作用定位聚合酶 I 的假设是一致的。聚合酶看起来与 DNA 接触，因为它扩展了 TIF-IB 形成的足迹，但这种接触是非特异性的。

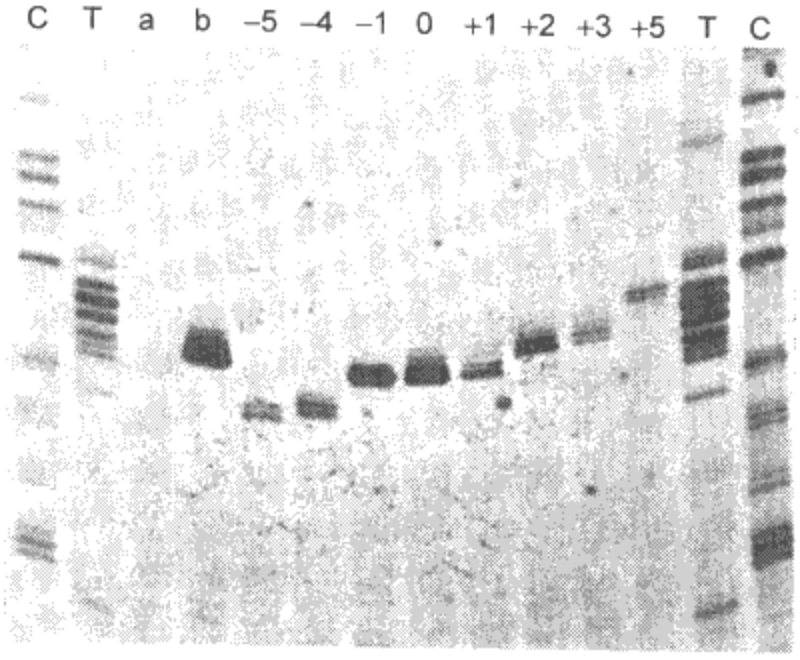


图 11.31　对聚合酶 I 转录起始位点插入和删除碱基的效应。Paule 及同事在卡氏棘阿米巴变形虫 rRNA 启动子的 TIF-IB 结合位点和正常转录位点间插入和删除多至 5 个碱基，如图上部所示。然后在体外转录这些模板，并用^{32}P 标记的 17 核苷酸测序引物进行引物延伸分析。将标记的延伸引物和用同样引物进行的 C 和 T 的测序产物（泳道 C 和 T）同时进行电泳。泳道 a，载体 DNA 的阴性对照，没有 rRNA 启动子；泳道 b，阳性对照，含野生型 rRNA 启动子；泳道 0，未删除的野生型 DNA 转录物的引物延伸。（*Source*：Reprinted from *Cell* v. 50，Kownin et al.，p. 695 © 2001，with permission from Elsevier Science.）

小结　I 类启动子被两个转录因子识别，核心结合因子和 UPE 结合因子。人的核心结合因子称为 SL1，在其他生物中，如卡氏棘阿米巴变形虫，与之同源的因子称为 TIF-IB。核心结合因子是募集 RNA 聚合酶 I 的基本转录因子。该因子有物种特异性，至少在动物中是这样。在哺乳动物和大多数其他生物中，结合 UPE 的因子被称为 UBF，但在酵母中称为 UAF，是帮助核心结合因子结合到核心启动子元件上的装配因子。在不同生物中，对 UPE 结合因子依赖程度有显著不同。卡氏棘阿米巴变形虫的 TIF-IB 自身就可以募集 RNA 聚合酶 I，并将其正确定位以便起始转录。

UPE 结合因子

人类中的 SL1 因子本身不能直接结合到 rRNA 启动子上，但经过部分纯化的 RNA 聚合酶 I 制备物中的 SL1 却能够做到。于是，Tjian 及同事开始从中寻找 DNA 结合蛋白，终于在 1988 年获得了人 UBF 的纯化产物。纯化的 UBF 由 97kDa 和 94kDa 两个多肽构成，其中 97kDa 多肽自身就有 UBF 活性。用高纯度 UBF 进行足迹分析时发现它与部分纯化的聚合酶 I 制备物在启动子上行为相似，即在核心元件和 UPE 的部分区段（称为位点 A）上具有相同的足迹，而 SL1 可加强此足迹，并使它扩展至 UPE 的另一部分（称为位点 B）（图 11.32）。因此，以前实验中与启动子结合的是 UBF 而不是聚合酶 I，而 SL1 使结合变得更容易。该研究没有证实 SL1 在含 UBF 的复合物中是否与 DNA 接触，或者仅改变了 UBF 的构象使其与延伸至位点 B 的更长 DNA 接触。基于这一结果和其他结果，可以认为 SL1 自己不能结合 DNA，UBF 可以，但是 SL1 和 UBF 是协同性地结合，并由此产生比其各自单独结合更广的结合效应。

Tjian 及同事还发现 UBF 促进 rRNA 基因的体外转录。图 11.33 显示利用人野生型 rRNA 启动子和突变体启动子（$\Delta 5'$-57，缺失 UPE），在不同 SL1 和 UBF 组合条件下进行转录的结果。所有反应中都有聚合酶 I，通过 S1 作图技术分析转录效率。泳道 1 含 UBF 但无 SL1，显示两个模板都不转录，再次证明 SL1 是转录绝对必需的。泳道 2 含 SL1 但无 UBF，显示本底水平的转录，再次证明 SL1 自身能够激活基本的转录。再看下面的图，无 UPE 模板的转录水平不如野生型高，表明 UBF 需通过 UPE 来激活转录。泳道 3 和 4，在有 SL1 和渐增 UBF 的条件下明显加强了两个模板的转录，特别是含 UPE 的模板。于是他们得出 UBF 作为转录因子能够结合到 UPE 上激活转录，但无 UPE 时，UBF 也可能结合到核心元件上发挥作用。

小结 人类 UBF 是通过聚合酶 I 激活转录的转录因子，可激活完整启动子或单一的核心元件，通过 UPE 介导激活。UBF 与 SL1 协同促进转录。

SL1 的结构和功能

前面讨论了两种人的转录因子 UBF 和 SL1，它们参与聚合酶 I 的转录。其中，UBF 是 97kDa 的单链多肽。但本章介绍过，TBP 是 I 类基因转录所必需的。那么 TBP 属于哪类因子呢？1992 年，Tjian

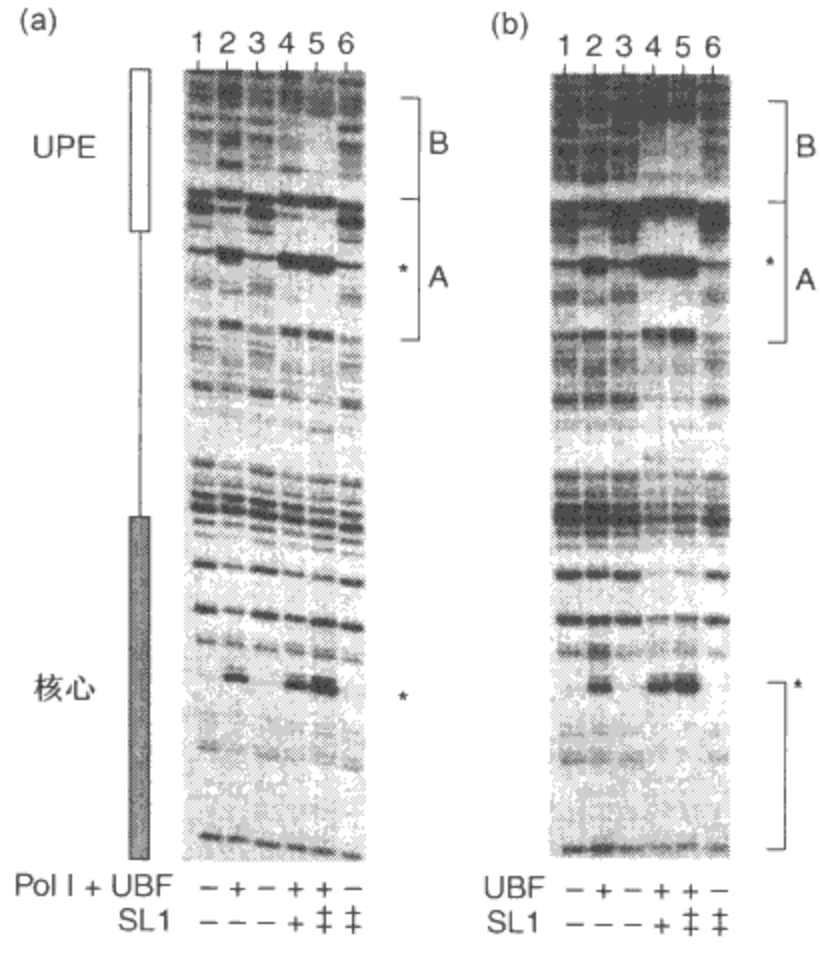
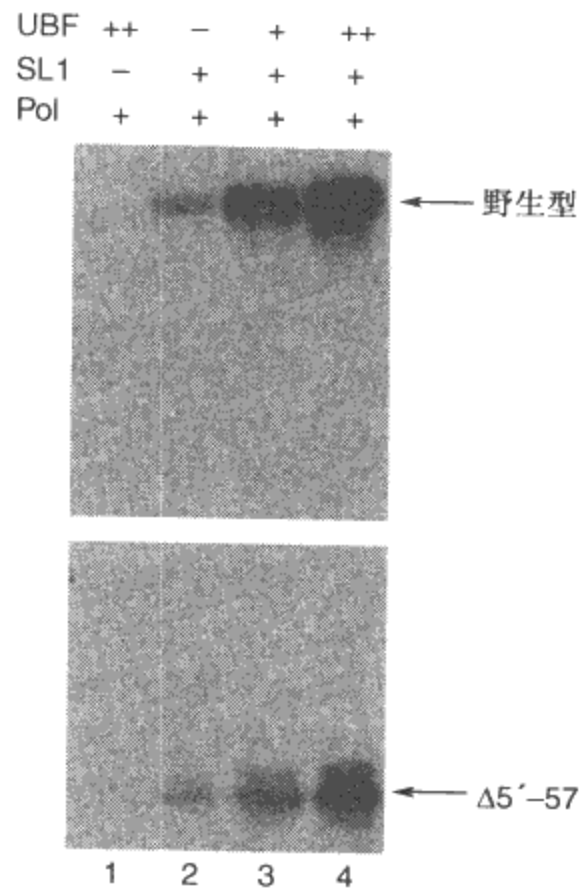


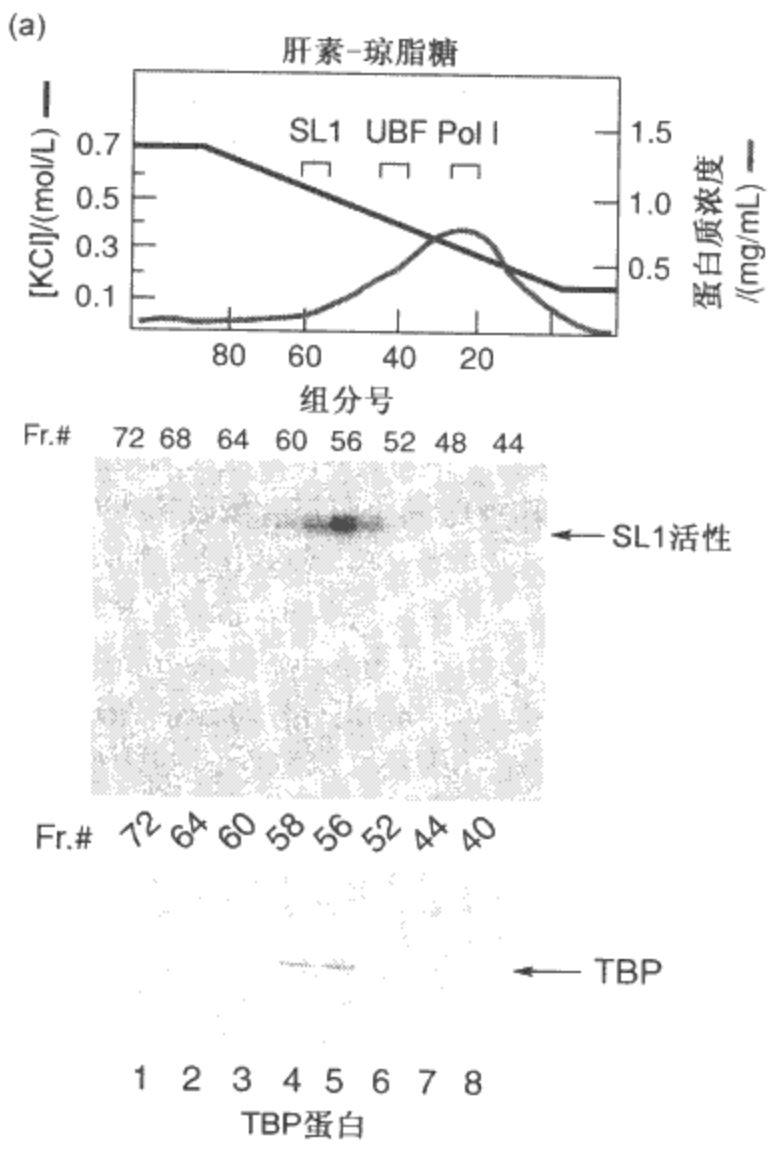
图 11.32 UBF 和 SL1 与 rRNA 启动子的相互作用。Tjian 及同事用人 rRNA 启动子与各种不同蛋白质组合进行 DNase I 足迹实验。(a) 聚合酶 I +UBF 与 SL1。(b) UBF 与 SL1。所用蛋白质在图底部标出。UPE 与核心元件的位置在左边标出，A 和 B 位点用括号在右边标出。星号表示 DNase 敏感性增强的位置。SL1 本身不能形成足迹，但对 UBF 在 UPE 和核心元件处的足迹可以增强并延伸。这种增强作用在无聚合酶 I 时尤为明显（b 图）。(*Source*: Adapted from Bell S. P., R. M. Learned, H. -M. Jantzen, and R. Tjian, Functional cooperativity between transcription factors UBF1 and SL1 mediates human ribosomal RNA synthesis. *Science* 241 (2 Sept 1988) p. 1194, f. 3 a-b.)

及同事证明：SL1 由 TBP 和 3 个 TAF 组成。首先，他们通过几种不同程序纯化了人（HeLa 细胞）的 SL1，每一步纯化之后都用 S1 分析法确定 SL1 的活性。然后，用 Western 印迹分析这些组分中 TBP 的存在。图 11.34 显示 SL1 的活性和 TBP 含量显著相关。

如果 SL1 确实包含 TBP，那么用抗 TBP 的抗体就可能抑制 SL1 的活性，Tjian 及同事证明确实如此。将细胞核提取物的 SL1 活性用抗 TBP 抗体耗尽，然后添加 SL1 可恢复其活性，但是只加入 TBP 却不能恢复 SL1 活性。所以，除了 TBP 以外，一定还有其他东西被移去了。

图 11.33　SL1 和 UBF 转录 rRNA 启动子的激活。在 RNA 聚合酶 I 和不同 UBF 和 SL1 组合条件下，Tjian 及同事用 S1 分析法测定人 rRNA 启动子的转录活性，如图顶部所示。上图显示野生型启动子的转录，下图显示无 UPE 功能的突变启动子（Δ5′-57）的转录。SL1 对本底转录是必需的，而 UBF 对两种模板的转录均起增强作用。(*Source*: Bell S. P., R. M. Learned, H.-M. Jantzen, and R. Tjian, Functional cooperativity between transcription factors UBF1 and SL1 mediates human ribosomal RNA synthesis. *Science* 241 (2 Sept 1988) p. 1194, f. 4. Copyright © AAAS.)





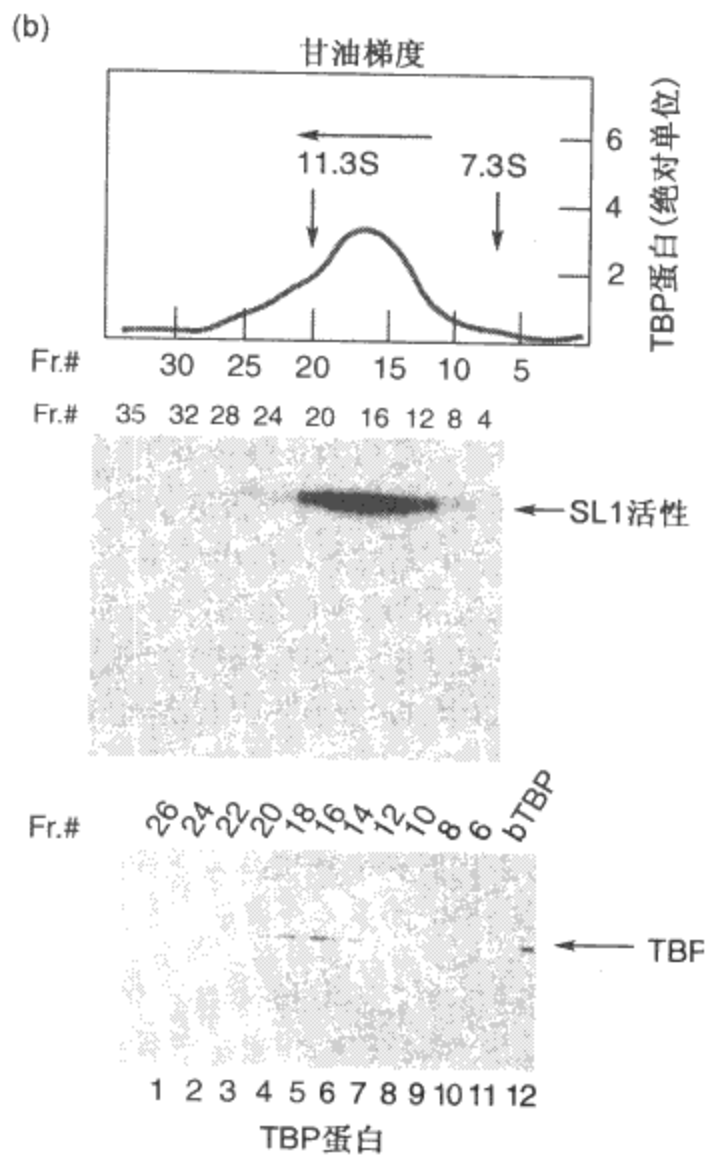


图 11.34　SL1 和 TBP 的共纯化。(a) 肝素-琼脂糖柱层析（见第 5 章柱层析方法）。上图，总蛋白从柱中洗脱的模式（红色）、盐浓度（蓝色）及 3 种特定蛋白质（括号）。中图，通过 S1 保护实验对选定洗脱组分进行 SL1 活性分析。下图，用 Western 印迹检测选定组分中的 TBP 蛋白。SL1 和 TBP 都集中在第 56 组分里。(b) 甘油梯度超速离心。上图，TBP 沉淀曲线。沉降系数分别为 11.3S 和 7.3S 的过氧化氢酶和醛缩酶作为参照，与 TBP 同时离心。中图和下图与 (a) 中一致。SL1 和 TBP 都沉降到以组分 16 为中心的位置。(*Source*: Comai, L., N. Tanese, and R. Tjian, The TATA-binding protein and associated factors are integral components of the RNA polymerase I transcription factor, SL1. *Cell* 68 (6 Mar 1992) p. 968, f. 2a-b. Reprinted by permission of Elsevier Science.)

在免疫共沉淀过程中还有哪些其他因子与 TBP 同时移去了呢？为此，Tjian 及同事对免疫沉淀物进行 SDS-PAGE。图 11.35 显示实验结果。沉淀物中，除了 TBP 和抗体（IgG）外，还有 3 种多肽，其分子质量分别是 110kDa、63kDa 及 48kDa（虽然 48kDa 条带因与 TBP 条带在一起而不易分辨）。因为这些多肽是与 TBP 一起免疫沉淀的，它们一定是 TBP 紧密结合的成分，因此属于 TBP 相关因子（TAF_I）。Tjian 把这些多肽称为 TAF_I110、TAF_I63 和 TAF_I48。它们与 TFIID 中的 TAF 完全不同（比较泳道 4 和 5）。用 1mol/L 的盐酸胍处理免疫沉淀物并再次沉淀后，可从 TBP 和抗体上剥离这些 TAF。抗体和 TBP 一起留在沉淀中（泳道 6）而各 TAF 留在上清中（泳道 7）。将纯化的 TBP 和 3 种 TAF 混合可以重建 SL1 活性，而且正如我们所期望的，这种活性具有物种特异性。在后来的工作中，Tjian 及合作者证明 TAF_I 和 TAF_{II} 可以竞争结合 TBP。这一发现提示这两组不同的 TAF 在 TBP 上的结合是相互排斥的。

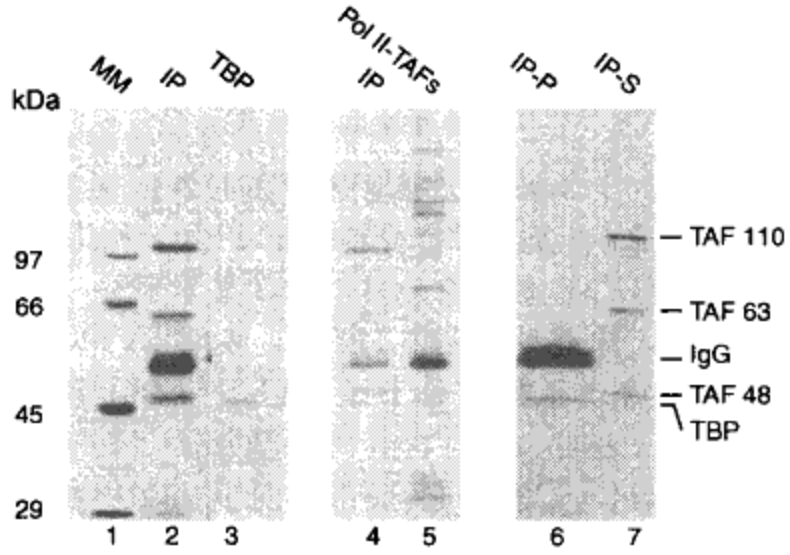


图 11.35　SL1 中的不同 TAF。Tjian 及同事用抗 TBP 的抗体对 SL1 进行免疫共沉淀，并对沉淀物的多肽进行 SDS-PAGE。泳道 1，分子质量标记；泳道 2，免疫沉淀物；泳道 3，用以比较的纯 TBP；泳道 4，另一种免疫沉淀样品；泳道 5，用于比较的 TFIID TAF（PolII-TAF）；泳道 6，用 1mol/L 盐酸胍处理免疫沉淀物再沉淀后的组分，显示 TBP 和抗体（IgG）；泳道 7，用 1mol/L 盐酸胍处理免疫沉淀物再沉淀后的上清，显示 3 种 TAF（在右侧标出）。（*Source*: Comai, L., N. Tanese, and R. Tjian, The TATA-binding protein and associated factors are integral components of the RNA polymerase I transcription factor, SL1. *Cell* 68 (6 Mar 1992) p. 971, f. 5. Reprinted by permission of Elsevier Science.）

因此，聚合酶 I 和聚合酶 II 的转录都依赖于 TBP 和几种 TAF 组成的转录因子（SL1 和 TFIID）。在两种因子中 TBP 是相同的，但 TAF 完全不同。

除酵母外，所有 II 类核心结合因子都是 TBP。而酵母 TBP 以不稳定状态结合到核心结合因子上，其方式与其他 TBP 结合至相应 TAF_I 的方式一致。我们在这里所讨论的 TAF_I 的数量和大小是人细胞所特有的。其他生物有各自的 TAF_I。

小结　人 SL1 由 TBP 和 3 种 TAF_I 组成，TAF_I110、TAF_I63 和 TAF_I48。纯化的各组分混合可重建全功能及具有物种特异性的 SL1。TBP 与 TAF_I 的结合阻止了 TBP 与 TAF_{II} 的结合。其他生物也有各自的一组 TAF_I。

11.3　III 类转录因子

1980 年，Roeder 及同事发现了可以结合到 5S rRNA 基因内部启动子并激活其转录的因子，并将其命名为 **TFIIIA**。此后另外两个因子 **TFIIIB** 和 **TFIIIC** 也被发现。这两个因子不仅参与 5S rRNA 基因的转录，而且参与聚合酶 III 的所有转录。

Barry Honda 和 Robert Roeder 在以非洲爪蟾为材料，首次建立真核生物体外转录系统时，证明了 TFIIIA 在 5S rRNA 基因转录中的重要性。他们发现，系统中加入 TFIIIA 后才能合成 5S rRNA。接着，Donald Brown 及同事证明当 5S rRNA 基因和 tRNA 基因加入到类似的无细胞提取物中时，可同时产生 5S rRNA 和 tRNA。而且，抗 TFIIIA 的抗体可以有效阻止 5S rRNA 的生成，但对 tRNA 的合成没有影响（图 11.36）。因此，TFIIIA 为 5S rRNA 基因转录所必需，但对 tRNA 则不然。

如果 tRNA 基因的转录不需要 TFIIIA，那么它需要什么因子呢？1982 年，Roeder 及同事分离出两个新的因子并称之为 **TFIIIB** 和 **TFIIIC**。他们发现这两个因子对 tRNA 基因转录是必需且充分的。随后知道这两种因子调控包括 5S rRNA 基因在内的所有聚合酶 III 基因的转录。这意味着前面提到的只需加入 TFIIIA 就能产生 5S rRNA 的细胞提取液，肯定含有 TFIIIB 和 TFIIIC。

小结　所有典型 III 类基因的转录都需要 TFIIIB 和 TFIIIC。5S rRNA 基因的转录需要这两个因子再加 TFIIIA。

TFIIIA

作为第一个真核转录因子，TFIIIA 的发现引

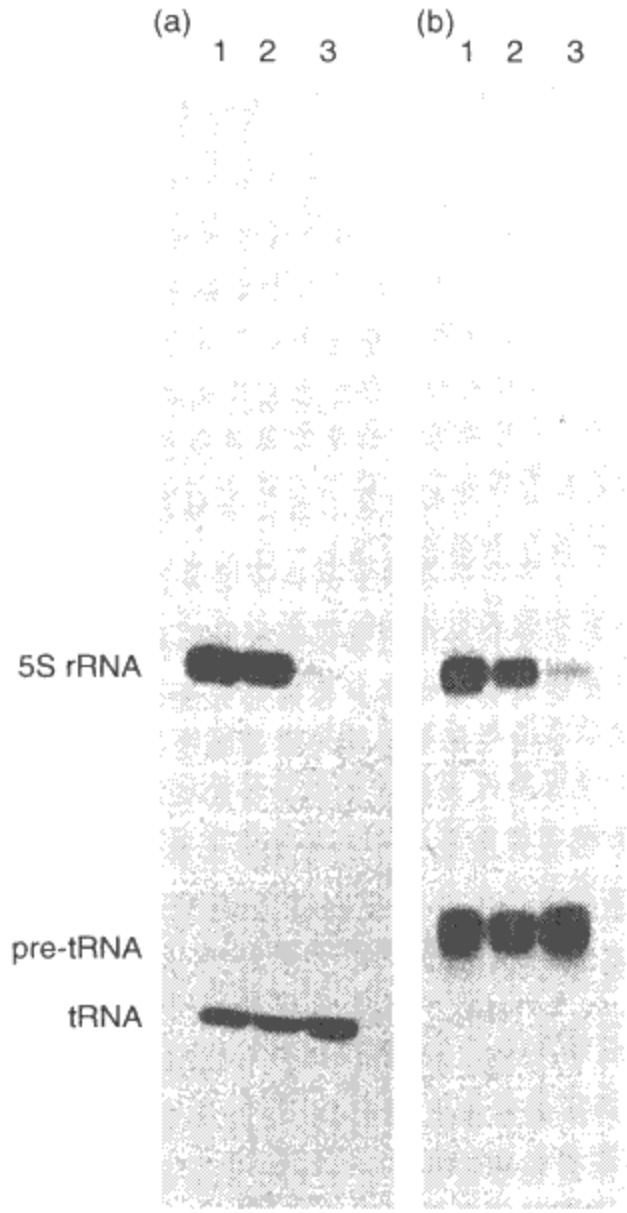


图 11.36　抗 TFIIIA 对聚合酶 III 转录的影响。Brown 及同事把克隆的 5S rRNA 和 tRNA 基因分别加入不同的细胞提取物中，并在标记的核苷酸及不同抗体条件下进行转录。(a) 卵母细胞提取物。(b) 体细胞提取物。泳道 1，无抗体；泳道 2 含有无关抗体；泳道 3 含有抗 TFIIIA 抗体。对转录后标记的 RNA 进行电泳，在两种提取物中，抗 TFIIIA 抗体都阻断了 5S rRNA 基因的转录，但不影响 tRNA 基因的转录。卵母细胞提取物可将 tRNA 前体加工为成熟 tRNA，但是体细胞提取物却不能。此外，在两种情况下转录都能发生。(*Source*: Pelham, H. B., W. M. Washington, and D. D. Brown, Related 5S rRNA transcription factors in *Xenopus* oocytes and somatic cells. *Proceedings of The National Academy of Sciences USA* 78 (Mar 1981) p. 1762, f. 3.)

起了极大的关注。它是 DNA 结合蛋白家族中所谓锌指大类的第一个成员，第 12 章将详细讨论锌指蛋白，这里我们先来看看 TFIIIA 的锌指。锌指的实质是一个粗略指形蛋白域，含有结合在一个锌离子周围的 4 个氨基酸。在 TFIIIA 和其他典型的锌指蛋白中，4 个氨基酸由 2 个半胱氨酸和 2 个组氨酸组成。而其他一些指形（finger-like）蛋白只有 4 个半胱氨酸，没有组氨酸。TFIIIA 的 9 个锌指排成一行，可以插入 5S rRNA 基因内部启动子任一边的大沟内，使特定的氨基酸与特定碱基对相互作用，形成一个紧密的蛋白质-DNA 复合物。

TFIIIB 和 TFIIIC

TFIIIB 和 TFIIIC 都是典型聚合酶 III 基因转录所必需的。因为它们的活性相互依赖，所以很难分开讨论。1989 年，Peter Geiduschek 及同事获得了一个粗制的转录因子制备物，它结合着 tRNA 基因的内部启动子及上游区域。图 11.37 中的 DNase 足迹实验证明了这一制备物的情况。泳道 c 未添加任何蛋白质。泳道 a 添加了各种转录因子和聚合酶 III。泳道 b 添加 a 中的各种蛋白质再加 3 种核苷酸（ATP、CTP 和 UTP），使转录进行 17nt，直至需要第一个 GTP。注意泳道 a 中，各种转录因子和聚合酶都能强烈保护内部启动子的 B 盒和上游区（U），但微弱保护内部启动子的 A 盒。泳道 b 显示聚合酶向下移动使得与 A 盒重叠的一个新区域被保护起来。但是聚合酶移走后，对上游区域的保护作用仍然存在。

那么，什么东西稳定地结合在上游区呢？为此，Geiduschek 及同事部分纯化了 TFIIIB 和 TFIIIC，并进行了足迹实验。图 11.38 显示其中一个实验结果。泳道 b，只有 TFIIIC，表明 TFIIIC 保护了内部启动子特别是 B 盒，但不结合上游区域。泳道 c 表明，同时含有两个转录因子时，上游区受到保护。类似的实验都证明了 TFIIIB 自身不能结合其中任何一个区域，它的结合完全依赖于 TFIIIC。然而，一旦 TFIIIC 帮助 TFIIIB 结合到上游区后，TFIIIB 就停留在那里，甚至聚合酶向前移走了也不动（回顾图 11.37）。图 11.38 的泳道 d 表明，即便 TFIIIC 从内部启动子上被肝素剥离，TFIIIB 仍与 DNA 结合，因为当 A 盒和 B 盒不再受保护时，上游区域仍然被保护着不受 DNase 降解。

根据目前的证据提出转录因子参与聚合酶 III 转录的以下作用模式（图 11.39）。首先，TFIIIC（或 TFIIIA 和 TFIIIC，对 5S rRNA 基因而言）结合到内部启动子上，接着这些装配因子帮助 TFIIIB 结合到上游区，而后 TFIIIB 帮助聚合酶 III 结合到转录起始位点，最后聚合酶转录基因。在此过程 TFIIIC（或 TFIIIA 和 TFIIIC）可能被除去，但 TFIIIB 保持结合，以促进后续转录。

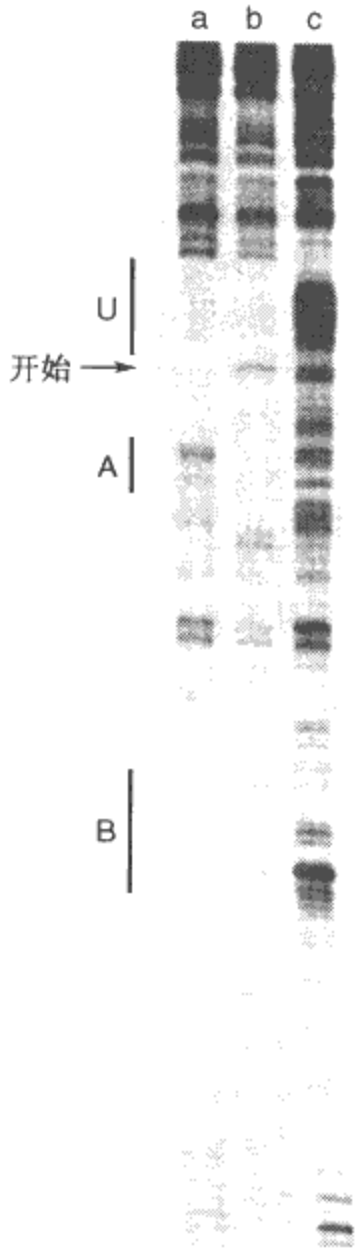


图 11.37　转录因子与 tRNA 基因的结合对转录的影响。Geiduschek 及同事用 tRNA 基因与含有聚合酶 III、TFIIIB 和 TFIIIC 的提取物进行了 DNase 足迹实验。泳道 a 含有转录因子，无核苷酸；泳道 b 含有转录因子和 3 种核苷酸（除 GTP），转录只能进行 17nt，此处需要第一个 GTP；泳道 c 为无蛋白质对照。泳道 b 中，聚合酶相对于泳道 a 的 17bp 移动使起始位点周围的足迹向下游移动，盖住 A 盒。紧邻起始位点上游区的足迹仍保持不变。（*Source*：Kassavetis，G. A.，D. L. Riggs，R. Negri，L. H. Nguyen，and E. P. Geiduschek，Transcription factor III B generates extended DNA interactions in RNA polymerase III transcription complexes on tRNA genes. *Molecular and Cellular Biology*. 9，no. 171（June 1989）p. 2555，f. 3. Copyright © 1989 American Society for Microbiology，Washington，DC. Reprinted with permission.）

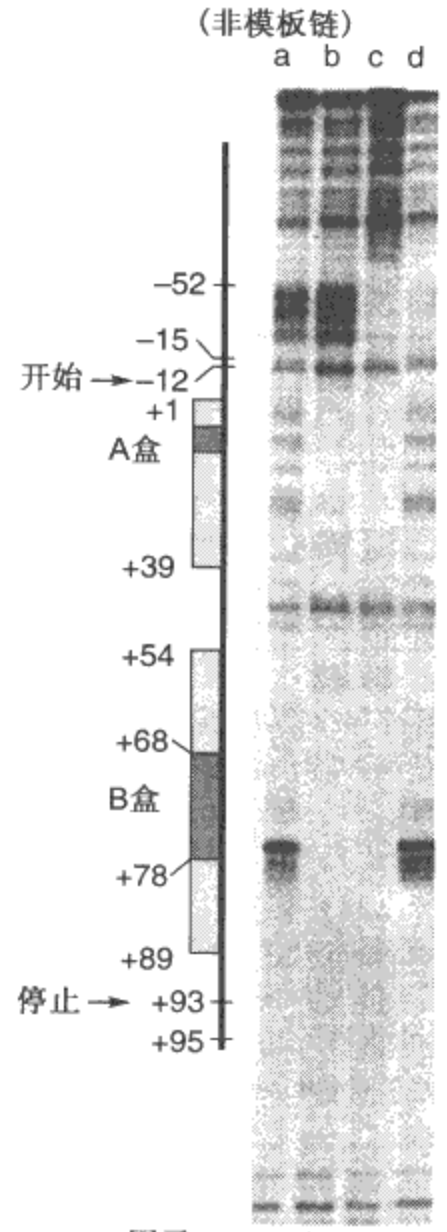


图 11.38　TFIIIB、TFIIIC 与 tRNA 基因的结合。Geiduschek 及同事用标记的 tRNA 基因（所有泳道）和纯化 TFIIIB 和 TFIIIC 的不同组合进行 DNase 足迹实验。泳道 a 无转录因子的阴性对照；泳道 b 仅有 TFIIIC。泳道 c，TFIIIB 加 TFIIIC；泳道 d，先加入 TFIIIB 和 TFIIIC，再加入肝素以除去松散结合的蛋白质。注意在 TFIIIC 基础上加入 TFIIIB 对上游保护区域的扩展（泳道 c），还需注意由 TFIIIC 对上游区的这一保护不受肝素影响，但对 A 盒和 B 盒的保护却未能幸免。黄色条框表示成熟 tRNA 的编码区，其中的 A 盒和 B 盒以蓝色表示。（*Source*：From Kassavetis，G. A.，D. L. Riggs，R. Negri，L. H. Nguyen，and E. P. Geiduschek，Transcription factor III B generates extended DNA interactions in RNA polymerase III transcription complexes on tRNA genes. *Molecular and Cellular Biology* 9：2558，1989. Copyright © 1989 American Society for Microbiology，Washington，DC. Repr-inted by permission.）

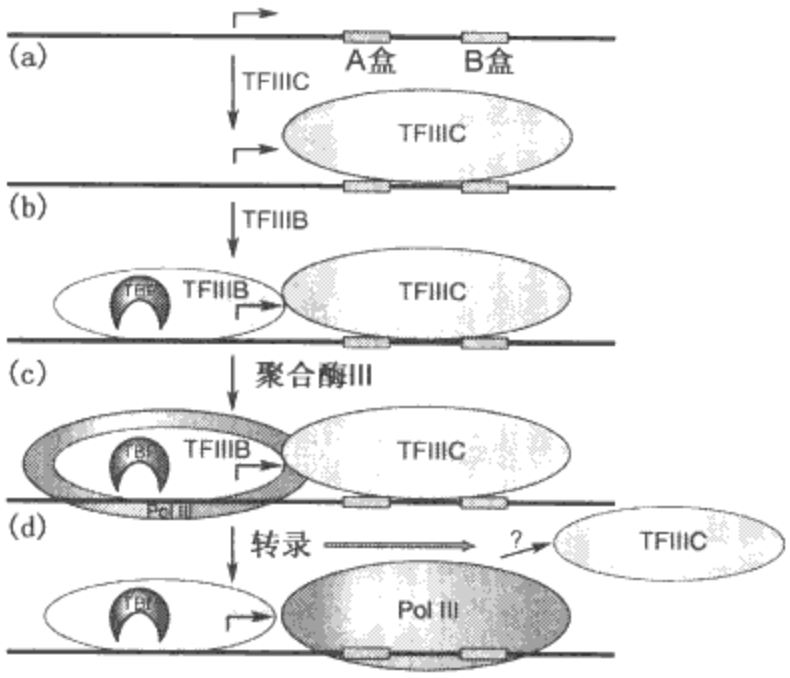


图 11.39　经典聚合酶 III 启动子（tRNA 基因）上前起始复合物的装配和转录起始的假设模式图。(a) TFIIIC（浅绿色）与内部启动子的 A 盒、B 盒（绿色）结合；(b) TFIIIC 促使 TFIIIB（黄色）以其 TBP（蓝色）结合转录起始位点的上游区；(c) TFIIIB 促使聚合酶 III（红色）结合到起始位点，准备起始转录；(d) 转录开始。聚合酶向右移动，产生 RNA（图中未显示）。此过程中 TFIIIC 可能被除去，也可能不被除去，但 TFIIIB 仍结合在原处准备起始下一轮聚合酶结合与转录。

Geiduschek 及同事提供了进一步的证据支持这个假说。让 TFIIIC 和 TFIIIB 与 tRNA 基因结合（或 TFIIIA、TFIIIC、TFIIIB 与 5S rRNA 基因结合），然后用肝素或高盐处理去除（剥离）装配因子 TFIIIC（或 TFIIIA 和 TFIIIC），接着将 TFIIIB-DNA 复合物与其他因子分离。最后，证明 TFIIIB-DNA 复合物仍能支持聚合酶 III 起始一至多轮转录（图 11.40）。TFIIIB 本身对启动子 DNA 并无亲和性，但它如何牢固地结合于靶 DNA 区呢？答案可能在于 TFIIIC（或 TFIIIA 和 TFIIIC）引起 TFIIIB 的构象变化，暴露出某个位点使 TFIIIB 能够牢固地与 DNA 结合。

TFIIIC 是一种不寻常的蛋白质，可结合 tRNA 基因的 A 盒和 B 盒，这已由 DNase 足迹和蛋白质-DNA 交联实验证明。有些 tRNA 基因在 A 盒和 B 盒之间存在内含子，但 TFIIIC 仍能设法结合到这两个启动子元件上。它是如何做到的呢？考虑 TFIIIC 是已知转录因子中最大且最复杂的，可能有助于理解这一问题。酵母 TFIIIC 含 6 个亚基，总分子质量约为 600kDa。电镜研究表明 TFIIIC 具有哑铃形结构，两个球状区域被一个可伸缩的接头区隔开，使整个蛋白质能跨越相当长的距离。

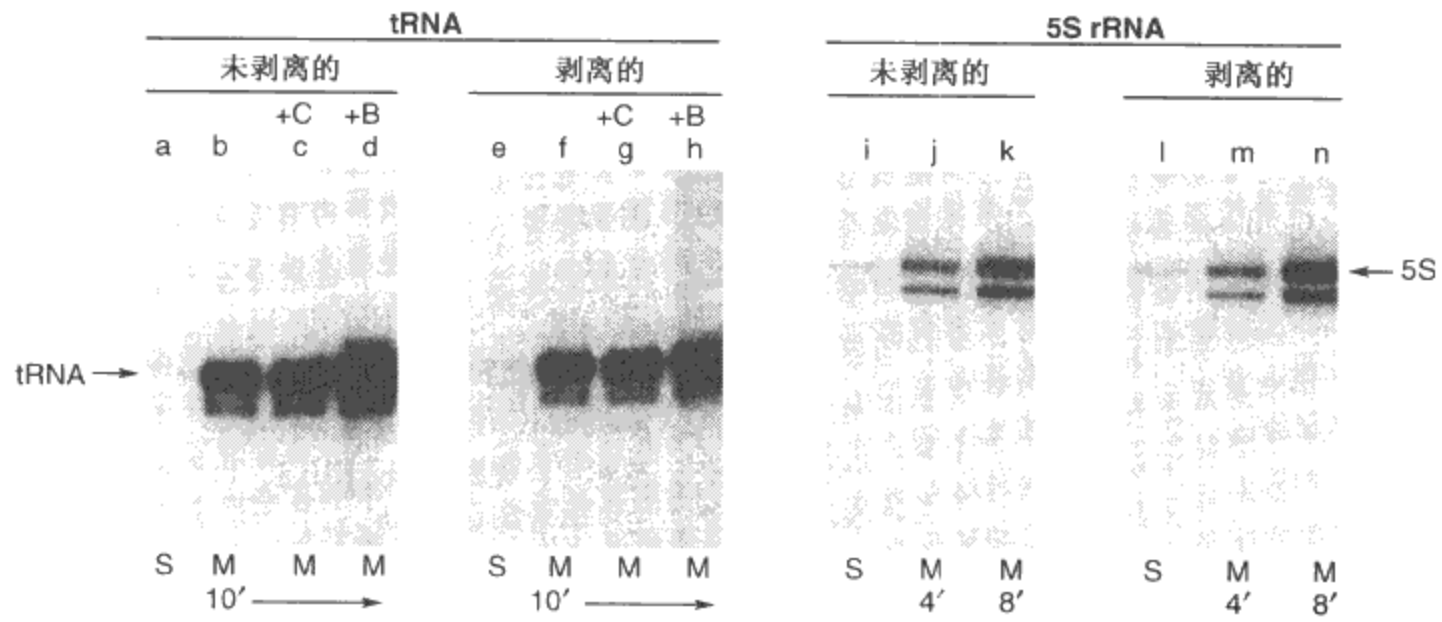


图 11.40　仅与 TFIIIB 结合的聚合酶 III 基因复合物的转录。Geiduschek 及同事获得了含有 tRNA 基因和 TFIIIB 及 TFIIIC 的复合物（左边两图），或 5S rRNA 与 TFIIIA、TFIIIB、TFIIIC 的复合物（右边两图）。用肝素除去 TFIIIC（泳道 e～h），或用高离子强度缓冲液除去 TFIIIA、TFIIIC（泳道 l～n）。把上述处理的模板经凝胶过滤柱除去未结合的因子。凝胶电泳和 DNase 足迹已证明（此处未显示）仅有 TFIIIB 结合在各基因的上游区，接着检测剥离处理的和未剥离处理的复合体起始单循环转录（S：泳道 a、e、i、l）和多循环转录（M：所有其他泳道）的能力，处理时间在底部标出。泳道 c 和 g 额外添加了 TFIIIC，泳道 d 和 h 额外添加了 TFIIIB，如图顶部所示。加入低浓度肝素，使泳道 a、e、i、l 中的转录仅进行一个循环并完成 RNA 的延伸，肝素限制了释放的聚合酶再起始转录。注意去除其他因子只保留 TFIIIB 的模板，无论在单循环还是多循环转录中，甚至当额外添加了 TFIIIC 时（比较泳道 c 和 g、k 和 n），其转录效率与未处理的模板差不多。未处理模板仅在 d 泳道效果比对应处理组的好，这是额外添加 TFIIIB 的结果，推测是由于剩余的游离 TFIIIC 帮助 TFIIIB 结合并形成了更多的前起始复合物。（*Source*: Kassavetis, G. A., B. R. Brawn, L. H. Nguyen, and E. P. Geiduschek, *S. cerevisiae* TFIIIB is the transcription initiation factor proper of RNA polymerase III, while TFIIIA and TFIIIC are assembly factors. *Cell* 60 (26 Jan 1990) p. 237, f. 3. Reprinted by permission of Elsevier Science.)

André Sentenac 及同事把酵母的 TFIIIC（他们又称其为 τ 因子）结合到一系列克隆的 tRNA 基因上，这些 tRNA 基因内 A 盒和 B 盒的间距不同，然后用透射电镜观察复合体的结构。图 11.41 显示当 A 盒和 B 盒间距为零时，TFIIIC 在 DNA 上呈现一个大斑点。随着间距增加，TFIIIC 逐渐呈现出两个球状区，中间被逐渐加长的连接区隔开。因此，大体积加上伸缩力，使 TFIIIC 能以其两个球状域与两个相距很远的启动子区结合。

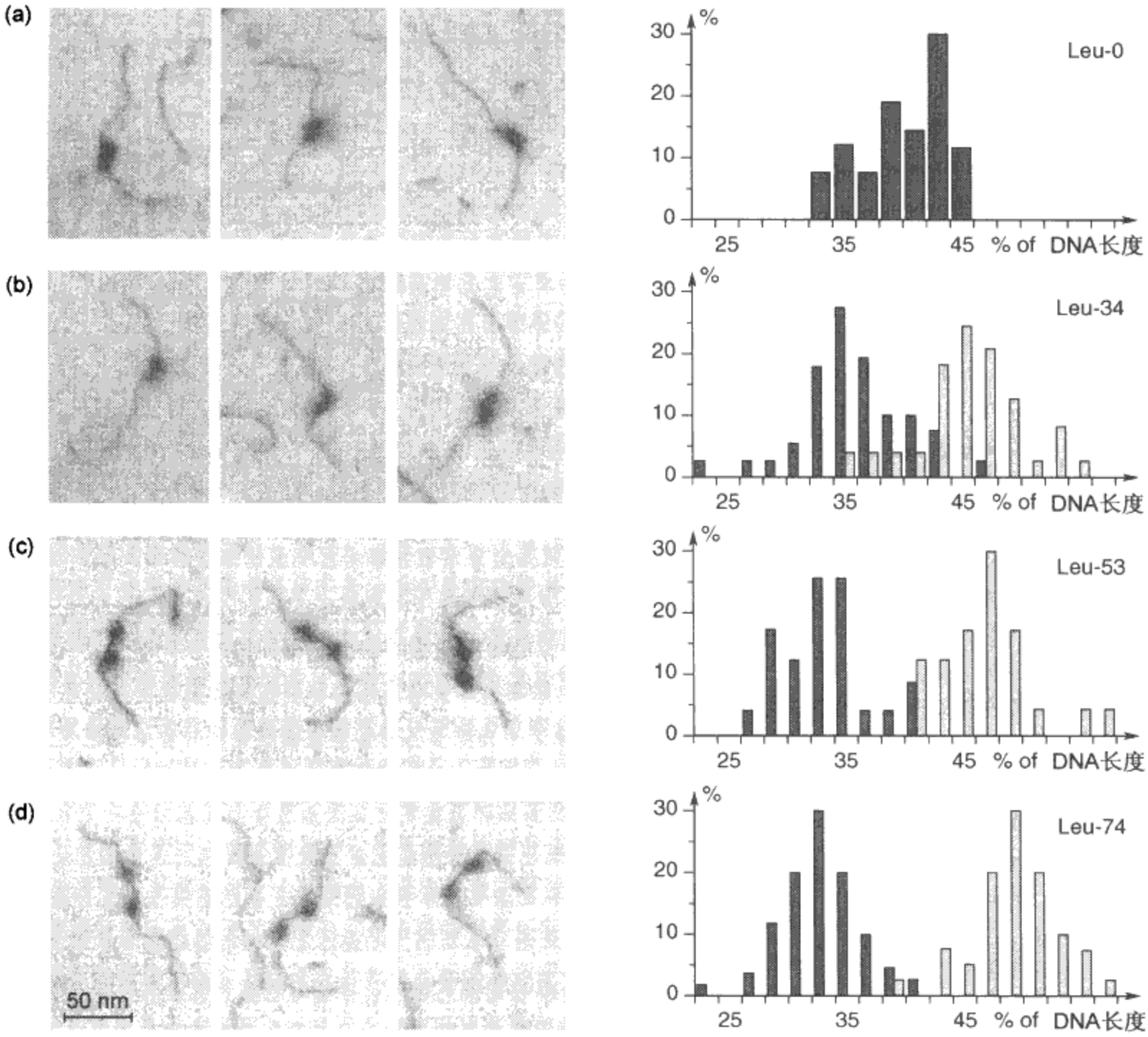


图 11.41　酵母 TFIIIC 含有两个球状域并由一个可伸缩区连接。Sentenac 及同事把酵母 TFIIIC 结合到一系列克隆的 tRNA 基因上，这些 tRNA 基因内 A 盒和 B 盒的间距不同。经醋酸双氧铀对复合物进行负染后用透射电镜观察。A 盒和 B 盒间的距离标在右侧：(a) 0bp；(b) 34bp；(c) 53bp；(d) 74bp（野生型的距离）。左边显示每种 DNA 的 3 张图像。右边柱状图显示 TFIIIC 球状结构在 DNA 上的位置，是大量显微图片的统计结果。柱子的高低表示沿 DNA 链的不同位点上球状结构域所占的百分比。红色柱表示靠近 DNA 末端的球状结构域，黄色柱表示另一个球状结构域的位置。（*Source*：Schultz et al *EMBO Journal* 8：p. 3817 © 1989.）

小结　为了与聚合酶形成前起始复合物，III 类基因需要 TFIIIB 和 TFIIIC 两个因子。5S rRNA 基因还需要 TFIIIA。TFIIIC 和 TFIIIA 结合到内部启动子区并协助 TFIIIB 结合到转录起始位点的上游，随后 TFIIIB 持续地结合在原处并不断地起始转录。TFIIIC 是一类分子质量很大的蛋白质，酵母中的此类蛋白质有 6 个亚基，形成两个球状区域并由一可伸缩的接头区所连接。接头的可伸缩性使它能覆盖内部启动子相距很远的 A 盒和 B 盒。

TBP 的作用

如果 TFIIIC 对 TFIIIB 结合经典 III 类基因是必需的，那么对于非经典 III 类基因没有 A 盒和 B 盒让 TFIIIC 结合，情况又如何呢？是什么促进 TFIIIB 对这类基因的结合？由于非经典 III 类基因的启动子含有 TATA 盒，并且已经了解这种启动子的转录需要 TBP，因此推测 TBP 结合到 TATA 盒上并把 TFIIIB 锚定在上游结合位点。

对于无 TATA 盒的经典聚合酶 III 类基因怎么办？已经知道 TBP 对经典 III 类基因的转录是必需的，如酵母和人类的 tRNA 基因和 5S rRNA 基因。这种情况下 TBP 结合在什么部位？现在已经清楚：TFIIIB 含有 TBP 及少量 TAF 因子。Geiduschek 及同事发现即使在最纯的 TFIIIB 里仍然有 TBP。对酵母 TFIIIB 的进一步研究，包括对其各克隆组分的重建实验，都揭示 TFIIIB 由 TBP 和 2 个 TAF_{III} 构成。这两种蛋白质在不同有机体中有不同命名，在酵母中被称作 B″和 TFIIB 相关因子（TFIIB-related factor）或 BRF，因其与 TFIIB 同源。

随后 Tjian 及同事通过因子回加至免疫耗尽（这些因子）的核提取物里，证实 TRFI 而非 TBP 对果蝇的 tRNA 基因、5S rRNA 基因和 U6 snRNA 基因转录是必需的，因此相对于其他生物对 TBP 依赖的普遍性，果蝇聚合酶 III 的转录是又一个例外。

对 3 种 RNA 聚合酶转录因子的研究揭示了一个一致规律：前起始复合物的形成始于装配因子，它能识别启动子内特定结合位点，然后由该蛋白质召集前起始复合物的其他组分。对于含 TATA 盒的 II 类启动子，其装配因子通常为 TBP，TBP 的结合位点是 TATA 盒。这也适用于含 TATA 盒的 III 类启动子，至少在酵母和人类细胞中是这样。我们已经熟悉在含 TATA 盒的 II 类启动子上这一过程如何开始的模型（图 11.4）。图 11.42 高度示意性地显示了所有无 TATA 盒启动子的前起始复合物的形成实质。I 类启动子的装配因子是 UBF，它首先结合到 UPE 上，然后吸引含有 TBP 的 SL1 结合到核心元件上。无 TATA 盒的 II 类启动子至少有两种方式吸引 TBP，TFIID 中的 TAF_{II} 可以结合到起始子上，或结合 Sp1，Sp1 已先结合到 GC 盒上。两种方式都可以将 TFIID 锚定到无 TATA 盒的启动子上。经典的 III 类启动子遵循同样的机制，至少在酵母和人类的细胞中是这样。TFIIIC 或 TFIIIA 和 TFIIIC（对 5S rRNA 基因）作为装配因子结合到内部启动子上，并吸引含 TBP 的 TFIIIB 结合到起始位点的上游。果蝇细胞中前起始复合物中的 TBP 被 TRFI 取代。

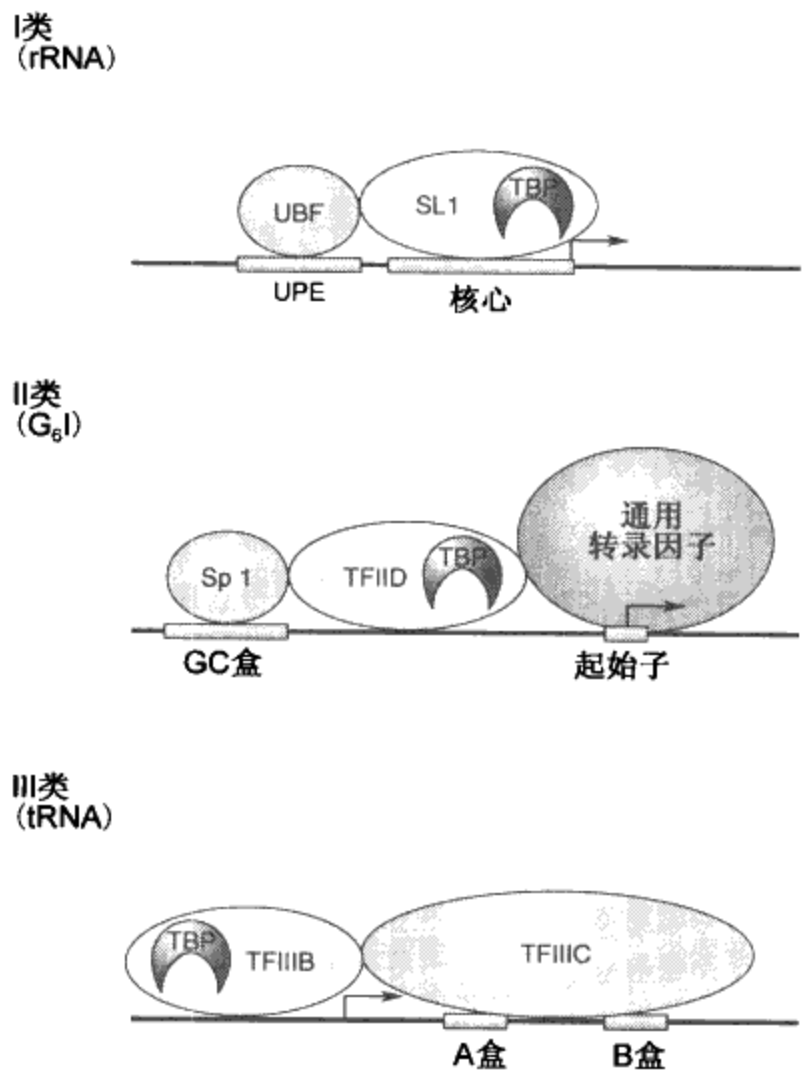


图 11.42　3 种聚合酶识别少 TATA 盒启动子前起始复合物的模型。每种情况下，装配因子（绿色）最先结合（在 I、II、III 类启动子上分别是 UBF、Sp1、TFIIIC）。然后吸引含 TBP（蓝色）的另一个因子（黄色），该因子在 I、II、III 类启动子分别是 SL1、TFIID、TFIIIB。对于 I、III 类启动子这些复合物足以结合聚合酶起始转录，但 II 类启动子除 RNA 聚合酶外还需更多通用因子（紫色）。（*Source*：Adapted from White, R. J. and S. P. Jackson, Mechanism of TATA-binding protein recruitment to a TATA-less class III promoter. *Cell* 71：1051，1992.）

不能因为 TBP 并不总是首先结合 DNA 而否认它在无 TATA 盒启动子中装配前起始复合物的重要性。一旦 TBP 结合上去，就包括 RNA 聚合酶在内的其他因子结合到复合物中。第二个一致规律：在大多数真核启动子中，TBP 在形成前起始复合物中起组织者作用。第三个一致规律是：TBP 的专一性由与之相伴的 TAF 决定，TBP 在与不同类型的启动子结合时，伴有不同类型的 TAF。

小结　在每一类真核启动子中，前起始复合物的装配都是从装配因子结合启动子开始的。在含 TATA 盒的 II 类启动子中（也可能是 III 类），此因子是 TBP，但其他类型的启动子有它们各自

的装配因子。即使在特定的启动子中，TBP 不是首先结合的装配因子，在大多数已知的启动子中它是前起始复合物的一部分，并在组建复合物过程中发挥组织作用。TBP 的特异性，即结合哪类启动子，是由与其相伴的 TAF 决定的。至少在果蝇 III 类基因的一些前起始复合物中，TRFI 代替了 TBP。

总结

各个转录因子在体外以下列顺序结合到 II 类启动子上（包括腺病毒主要晚期基因启动子）：①TFIID 在 TFIIA 的帮助下，结合 TATA 盒，形成 DA 复合物；②接着是 TFIIB 的结合；③TFIIF 帮助 RNA 聚合酶结合到－34bp 到＋17bp 之间的 DNA 区域，其余因子按 TFIIE、TFIIH 顺序结合，最后形成 DABPolFEH 前起始复合物。在体外，TFIIA 对复合物的形成是非必需的。

TFIID 含有一个 TATA 盒结合蛋白（TBP），外加至少 8 个多肽，即 TBP 相关因子（TAF_{II}）。人类 TBPC 末端的 180 个氨基酸是 TATA 盒结合域。TBP 和 TATA 盒在 DNA 小沟发生相互作用。马鞍形 TBP 与 DNA 呈直线排列，鞍的下部迫使小沟张开，使 TATA 盒呈 80°弯曲。TBP 是所有三类基因的大多数成员转录所必需的，并不只是对 II 类基因。

大多数 TAF_{II} 在真核生物进化中是保守的。TAF_{II}行使多种功能，两个显著的功能是与核心启动子元件和激活因子的相互作用。TAF_{II} 250 和 TAF_{II}150 有助于 TFIID 结合启动子的起始子和 DPE，使 TBP 与无 TATA 盒但有起始子和 DPE 的启动子结合。TAF_{II} 250 和 TAF_{II} 110 有助于 TFIID 与结合在转录起始位点上游 GC 盒上的 Sp1 相互作用，使 TBP 可以结合无 TATA 盒但有 GC 盒的启动子。与各种激活因子的作用，显然需要不同 TAF_{II} 的组合，至少在高等真核生物是这样。TAF_{II}250 还具有两种酶活性，即组蛋白乙酰转移酶活性和蛋白激酶活性。至少在高等真核生物，TFIID 并不是通用的。果蝇的某些启动子需要另一个因子 TRF1。而某些启动子需要不含 TBP 的 TAF_{II}复合物。

对 TFIIB-聚合酶 II 复合物的结构研究表明：TFIIB 通过 C 端域在 TATA 盒处结合 TBP，同时通过 N 端域结合聚合酶 II。这种桥梁作用使聚合酶活性中心粗定位于 TATA 盒下游 25～30bp 处。在哺乳动物中，TFIIB N 端域的环状基序发挥转录起始的精确定位，通过与很靠近活性中心的模板 DNA 单链间的作用而实现。生化研究证明在前起始复合物中，TFIIB 的 N 端域（具体在指域和接头区）位于靠近 RNA 聚合酶 II 活性中心及 TFIIF 最大亚基的位置。

前起始复合物形成过程中 RNA 聚合酶 II 以低磷酸化的 IIA 形式存在。随后 TFIIH 使 RNA 聚合酶 II 最大亚基的 C 端域（CTD）上的七氨基酸重复序列上的 2 位和 5 位丝氨酸磷酸化，由此产生聚合酶的磷酸化形式（IIO）。在体外 TFIIE 可极大地促进这一过程，磷酸化是转录起始所必需的。在由起始向延伸转变过程中，七氨基酸重复序列 5 位丝氨酸失去磷酸化。如果 2 位丝氨酸也失去磷酸化，聚合酶将暂停等待一个非 TFIIH 激酶进行磷酸化。

TFIIE 和 TFIIH 对形成开放启动子复合体或延伸并非必需，但在启动子清除中起作用。TFIIH 的解旋酶活性是转录必需的，推测是因为此活性使启动子区的 DNA 完全解链从而促进启动子清除。

转录可在延伸水平调控，TFIIS 因子通过限制聚合酶在不连续位点的转录停滞而刺激延伸。TFIIF 通过插入 RNA 聚合酶的活性中心，激发其核酸酶活性，切割排出的引起停滞的新生 RNA 的 3′端而发挥其功能。TFIIF 显然是通过限制瞬时停顿而激活延伸。TFIIS 也参与校正，通过激活 RNA 聚合酶中的 RNase 活性，切除错配的核苷酸。

酵母和哺乳动物细胞具有 RNA 聚合酶 II 全酶，除了聚合酶的亚基外还包括许多多肽，额外的多肽包括一系列通用转录因子（TBP 除外）和中介物。

I 类启动子被两个转录因子识别，核心结合因子和 UPE 结合因子。人类核心结合因子称为 SL1，在其他生物中，如卡氏棘阿米巴变形虫，与之同源的因子为 TIF-IB。核心结合因子是募集 RNA 聚合酶 I 的基本转录因子。该因子有物种特异性，至少在动物中是这样。在哺乳动物和大多数其他生物中，结合 UPE 的因子被称为 UBF，但在酵母中称为 UAF，是帮助核心结合因子结合到核心启动子元件上的装配因子。在不同生物中，对 UPE 结合因子依赖程度有显著不同。卡氏棘阿米巴变形虫的 TIF-IB 自身就可以募集 RNA 聚合酶 I，并将其正确定位以便起始转录。人 UBF 是通过聚合酶 I 激活转录的转录因子，可激活完整启动子或单一的核心元件，通过 UPE 介导激活。UBF 与 SL1 协同激活转录。

人 SL1 由 TBP 和 3 种 TAF_I 组成，即 TAF_I110、TAF_I63 和 TAF_I48。纯化的各组分混合可重建全功能及具有物种特异性的 SL1。TBP 与 TAF_I 的结合阻止了 TBP 与 TAFII 的结合。其他生物有自己的一组 TAF_I。

典型 III 类基因的转录都需要 TFIIIB 和 TFIIIC 形成与聚合酶的前起始复合物。5S rRNA 基因还需

要 TFIIIA。TFIIIC 和 TFIIIA 结合到内部启动子区并协助 TFIIIB 结合到转录起始位点的上游，随后 TFIIIB 持续地结合在原处并不断地起始转录。

在每一类真核启动子前起始复合物的装配都是从装配因子结合启动子开始的。有 TATA 盒的 II 类启动子（也可能是 III 类），该因子是 TBP，但其他类型的启动子有它们自己的装配因子。即使在特定的启动子中 TBP 不是首先结合的装配因子，但在大多数已知的启动子中它是前起始复合物的一部分，并在组建复合物过程中发挥组织作用。TBP 的特异性，即结合哪类启动子，是由与其相伴的 TAF 决定的，每类启动子的 TAF 都是专一的。

复习题

1. 列出在体外装配形成 II 类前起始复合物的蛋白质顺序。

2. 叙述实验并给出结果，显示 TFIID 是 II 类前起始复合物的基本组分。

3. 叙述实验并给出结果，显示 TFIIF 和聚合酶 II 结合，但二者均不能独自结合到前起始复合物上。

4. 叙述实验并给出结果，显示 TFIID 在复合物中的结合位点。

5. 给出 DAB 与 DABPolF 复合物的足迹差异，基于这种差异你能得出什么结论？

6. 提出一种假设解释在 TATA 盒中用 dC 替换 dT、dI 替换 dA（产生 CICI 盒）对 TFIID 的结合影响不大，给出假设的基本依据。

7. TBP 具有何种形状？TBP 与 TATA 盒作用的几何学是什么？

8. 叙述实验并给出结果，显示 TBP 对三类启动子的转录是必需的。

9. 叙述实验并给出结果，显示在体外结合 TFIID 的 II 类启动子比结合 TBP 时活性更强。

10. 叙述实验并给出结果，鉴定结合到含 TATA 盒、起始子及下游元件的 II 类启动子上的 TAF。

11. 描述 DNase 足迹实验并给出结果，显示与单独的 TBP 相比，$TAF_{II}250$ 和 $TAF_{II}150$ 使足迹扩大。

12. 图示 TBP（和其他因子）与少 TATA 盒的 II 类启动子的相互作用。

13. 全基因组表达分析表明 $TAF_{II}145$ 对酵母 16%的基因转录是必需的，$TAF_{II}17$ 对酵母 67%的基因转录是必需的。说明这些结果的基本原理。

14. 给出下列蛋白质与 II 类前起始复合物结合的例子：a. 可选择的 TBP；b. 丢失 TAF_{II}；c. 无 TBP 或 TBP 类似蛋白质。

15. 转录过程中 TFIIA 和 TFIIB 的显著作用是什么？

16. 图示 TBP-TFIIB-RNA 聚合酶 II 复合物与 DNA 的结合，显示蛋白质的相对位置。这些位置与蛋白质的功能有何关系？

17. 叙述实验并给出结果，绘出 Rpb1 和 Rpb2 上与 TFIIB 的指域和接头区紧密接触的位点。

18. 叙述实验并给出结果，显示 TFIIH 而非其他通用转录因子，使 RNA 聚合酶磷酸化由 IIA 型转变为 IIO 型，还要包括其他通用转录因子帮助 TFIIH 其磷酸化作用的结果。

19. 试述 TFIIH 使聚合酶 II 的 CTD 磷酸化的实验并给出结果。

20. 试述 DNA 解旋酶实验及如何用解旋酶证明 TFIIH 与解旋酶活性有关。

21. 叙述少 G 盒转录实验及如何用它证明在体外实验中与 TFIIH 相关的 RAD25 DNA 解旋酶活性是转录所必需的。

22. 绘出 II 类前起始复合物示意图，显示聚合酶、启动子 DNA、TBP 以及 TFIIB、TFIIE、TFIIF 和 TFIIH 的相对位置，并指出转录的方向。

23. 叙述实验并给出结果，显示 TFIIS 激活 RNA 聚合酶 II 的转录延伸。

24. 给出一个模型，显示 TFIIS 对转录停滞的逆转效应。TFIIS 的哪部分最直接参与此项任务？怎么参与？

25. 叙述实验并给出结果，显示 TFIIS 通过 RNA 聚合酶 II 激活校正。

26. RNA 聚合酶 II 全酶的概念是什么？RNA 聚合酶 II 全酶与 RNA 聚合酶 II 核心酶有何区别？

27. 叙述实验并给出结果，显示 I 类启动子的核心元件与转录起始位点之间增加或删除一些碱基的影响。

28. I 类启动子的哪个通用转录因子是装配因子？换句话说，哪个最先结合上去，并且帮助其他的因子结合？设计 DNase 足迹实验加以证明，并给出理想的结果（不一定与 Tjian 及同事的结果完全一致）。用图表示足迹中两转录因子的相互影响。

29. 叙述共纯化和免疫共沉淀实验，显示 SL1 包含 TBP。

30. 叙述在 SL1 中确定 TAF 存在的实验并给出结果。

31. 如何知道 TFIIIA 对 5S rRNA 基因而非 tRNA基因的转录是必要的？

32. Geiduschek 及同事用聚合酶 III、TFIIIB、TFIIIC 和 tRNA 基因进行 DNase 足迹实验。显示他们得到的下列组合的实验结果：不加入蛋白质；聚

合酶和转录因子；聚合酶、转录因子及 3 种 NTP。从中得出什么结论？

33. 经典 III 类启动子为内部启动子。TFIIIB 和 TFIIIC 一起在基因编码区上游区形成足迹，绘出这两种因子结合的示意图，并解释这些结果。

34. 做出聚合酶 III 已经开始转录经典 III 类基因（如 tRNA 基因）时，TFIIIB 和 TFIIIC 变化的示意图。该图如何解释新一轮的聚合酶 III 仍能继续转录，甚至当转录因子已不再与内部启动子结合时。

35. 叙述 DNase 足迹实验及其结果，显示 TFIIIB+TFIIIC，而非 TFIIIC 自身保护 tRNA 基因转录起始位点的上游区，并显示用肝素剥离 TFIIIC 后足迹的变化。

36. 叙述并给出结果，显示以下过程：一旦 TFIIIB 结合经典 III 类基因，它就可以支持多轮转录，甚至当 TFIIIC（或 TFIIIC 和 TFIIIA）已从启动子上去除了。

37. 叙述并给出结果，证明 TFIIIC 在结合 A 盒和 B 盒时的灵活性，无论 III 类启动子内 A 盒和 B 盒间距离远近。

38. 图示前起始复合物与所有三类少 TATA 盒启动子的结合。区分每种情况的组装因子。

分析题

1. 假如你正在研究一类新的真核生物启动子（IV 类），它由新的 RNA 聚合酶 IV 识别。你发现有两种通用转录因子是这类启动子转录所必需的。设计实验验证其中哪种是装配因子，哪种是募集 RNA 聚合酶结合启动子所必需的？给出预期的实验结果。

2. 你发现有一个 IV 类转录因子含有 TBP，设计实验鉴定这个因子含有 TAF。

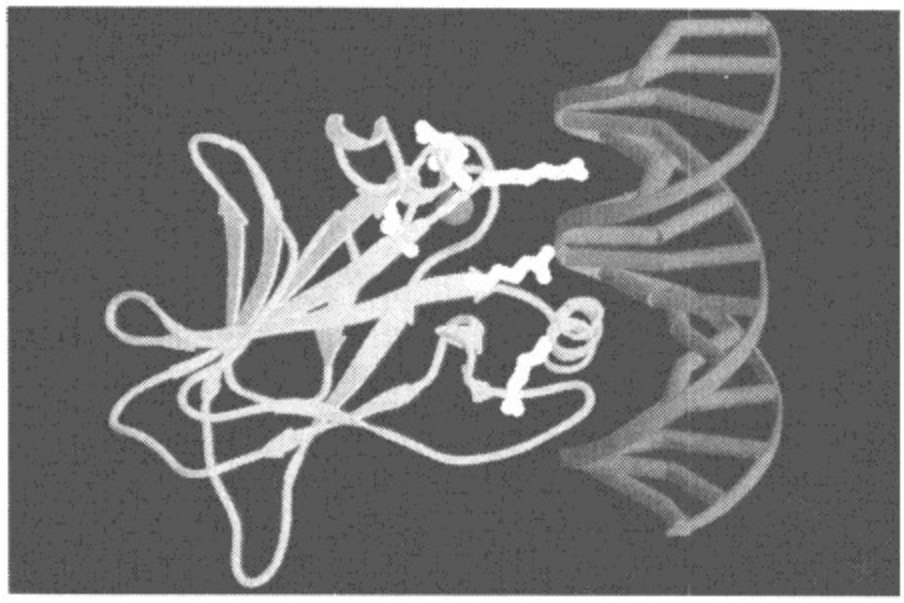
3. 一些 IV 类启动子含有两个 DNA 元件 X 盒和 Y 盒，另一些则只有 X 盒。设计实验鉴定与每种启动子结合的 TAF。

4. 将细胞与含有抑制 TFIIH 蛋白激酶活性的抑制剂温育，然后进行体外转录与 DNase 足迹实验。在转录的哪一步会受阻？什么实验能揭示出这种阻断作用？你是否仍期望在启动子区有一个足迹？为什么？如果是这样，与缺乏抑制剂的足迹比较，此实验中的足迹会有多大？

5. 已知蛋白质 X 和蛋白质 Y 相互作用，如何定位蛋白质 X 上与蛋白质 Y 作用的具体区域？设计羟基自由基切割实验回答此问题。

翻译　兰海燕　校译　马　纪

第 12 章　真核生物的转录激活因子



转录因子 p53 与 DNA 靶位点相互作用的计算机模型。

Courtesy Nicola P. Pavletich, Sloan-Kettering Cancer Center, Science (15 July 1994) cover. Copyright © AAAS

在第 10 章和第 11 章我们学习了真核生物转录的基本装置：3 种 RNA 聚合酶、对应的启动子以及将聚合酶与启动子联系起来的通用转录因子，但这些装置并不足以完成转录。通用转录因子虽然能够识别转录起始位点并指示转录方向，但它们自身只能激发很低的转录水平（本底水平的转录），而细胞中活跃基因的转录通常要高于本底转录。为了能够增强转录，真核细胞另有一类与增强子（一种 DNA 元件，第 10 章）结合的基因特异性转录因子（**激活因子，activator**），由激活因子产生的转录激活也使得细胞能够调控基因的表达。

此外，真核 DNA 与蛋白质形成一种称为染色质的结构。其中有些是异染色质，它们高度浓缩，RNA 聚合酶不能与之结合，所以不能转录。另一些是常染色质，同样含有蛋白质但结构相对伸展。虽然相对开放，多数常染色质所含基因在特定细胞中由于没有合适转录激活因子的开启也不能转录。甚至某些蛋白质会覆盖启动子使其不与 RNA 聚合酶和通用转录因子结合而保持关闭。本章将探讨调控真核基因转录的激活因子，在第 13 章将关注激活因子、染色质结构和基因活性之间的相互作用。

12.1　激活因子的类型

转录激活因子可以激活也可抑制 RNA 聚合酶 II 的转录。它们至少都有两个基本的功能域：**DNA 结合域**（DNA-binding domain）和**转录激活域**（transcription-activating domain）。很多激活因子还具有形成二聚体的域，能使两个单体彼此结合形成同源二聚体（两个相同单体彼此结合）、异源二聚体（两个不同单体彼此结合）或多聚体（如四聚体）。有些激活因子甚至还具有结合像类固醇激素这种效应分子的结合位点。下面我们讨论这 3 种功能域的实例。

DNA 结合域

蛋白质结构域是指蛋白质的独立折叠单位。每种 DNA 结合域都有一个 **DNA 结合模体**（DNA-binding motif），模体是结构域的一部分，具有特定的形状以结合特定的 DNA。大多数 DNA 结合模体可归为如下几类。

1. 含锌组件（zinc-containing module）。至少有三类含锌组件具有 DNA 结合模体的功能，它们利用一个或多个锌离子形成合适的构象，使 DNA 结合模体的 α 螺旋能进入 DNA 双螺旋大沟，并结合在特定的位置上。这类含锌组件包括：①锌指（Zinc finger），存在于前述的 TFIIIA 和 Sp1 转录因子中；②锌组件，在糖皮质激素受体和其他的细胞核受体中发现；③含 2 个锌离子和 6 个半胱氨酸组件，在酵母转录激活因子 GAL4 及其家族成员中发现。

2. 同源结构域（homeodomain，HD）。含有约 60 个氨基酸，与原核生物蛋白质中（如 λ 噬菌体阻遏物）螺旋-转角-螺旋的 DNA 结合域在结构和功能上类似。HD 最早发现于调控果蝇发育的激活因子同源盒蛋白

(homeobox protein)，广泛存在于各类激活因子中。

3. bZIP 和 bHLH 模体。CCAAT 盒/增强子结合蛋白（C/EBP）、MyoD 蛋白（肌细胞定向分化调控因子）以及其他真核细胞转录激活因子均有一个强碱性 DNA 结合模体，与 1 或 2 个蛋白质二聚化结构域（protein dimerization motif）连接，该结构域叫做亮氨酸拉链（leucine zipper）和螺旋-环-螺旋模体（helix-loop-helix，HLH）[C/EBP 与 CCAAT 结合转录因子（CTF，第 10 章）不同]。

以上所列并非详尽无遗。事实上，近来鉴定的好几种转录激活因子不能归入上述任何一类。

转录激活域

大多数激活因子只含有一种转录激活域，有些则有几种激活域。迄今为止，大多数的转录激活域可归为如下 3 种类型。

1. 酸性域（acidic domain）。酵母转录激活因子 GAL4 是典型代表。在由 49 个氨基酸组成的激活域中，有 11 个氨基酸为酸性氨基酸。

2. 富谷氨酰胺域（glutamine-rich domain）。Sp1 转录激活因子有两个这样的功能域，谷氨酰胺占该区氨基酸总数的 25%左右。其中一个功能域在 143 个氨基酸肽段中含有 39 个谷氨酰胺。此外，Sp1 还有两个转录激活域不能列入这三类的任何一种。

3. 富脯氨酸域（proline-rich domain）。如转录激活因子 CTF，在 84 个氨基酸组成的功能域中有 19 个是脯氨酸。

由于对转录激活域本身并不十分了解，所以对其描述只能是模糊的。例如，酸性域似乎只需要酸性氨基酸残基来发挥作用，故用“酸性团”（acid blob）来称呼这个似乎无结构的域。另一方面，Stephen Johnston 及同事已证实 GAL4 的酸性激活域在弱酸性溶液中趋向于形成一种精确的构象——β 折叠。在体内弱碱性条件下，可能也会生成 β 折叠，但这一点尚不明确。将 GAL4 酸性域中的所有 6 个酸性氨基酸除去后，该域仍保持 35%的转录激活能力。这样，不仅酸性转录激活域的结构不清楚，而且其酸性本质的重要性也值得怀疑了。

带着这一疑问，很难将转录激活域的结构和功能对应起来。此外，以后我们将看到 Sp1 富谷氨酰胺的转录激活域似乎要通过与其他转录激活因子的富谷氨酰胺区相互作用发挥功能。

小结 真核细胞的转录激活因子至少有两个域：DNA 结合域和转录激活域。DNA 结合域包含像含锌组件、同源结构域、亮氨酸拉链和螺旋-环-螺旋等模体。转录激活域有酸性域、富谷氨酰胺域和富脯氨酸域。

12.2 激活因子内 DNA 结合模体的结构

与转录激活域相比，DNA 结合域的结构研究得较清楚。X 射线晶体衍射实验显示了这种结构与靶基因之间的相互作用。此外，类似的结构分析实验已反复证明，二聚化结构域是促使蛋白质单体之间相互作用并最终形成功能性二聚体甚至四聚体的主要部分。这一点非常重要，因为 DNA 结合蛋白大多不能以单体形式结合靶基因序列，它们至少得形成二聚体才能发挥作用。我们来探讨几种 DNA 结合模体的结构，以及它们如何介导与 DNA 的相互作用。在此过程中我们将发现一些蛋白质分子二聚化的方式。

锌指结构

1985 年，Aaron Klug 注意到通用转录因子 TFIIIA 结构具有周期性的重复。由 30 个氨基酸残基组成的单元在蛋白质中重复了 9 次，每个重复序列由一对空间上彼此靠近的半胱氨酸紧随 12 个其他氨基酸，后接一对空间上彼此靠近的组氨酸构成。更重要的是这种蛋白质富含锌，每个重复单元有一个锌离子。Klug 由此预测锌指结构的共同特征是，在每个重复单元中通过一对半胱氨酸和一对组氨酸与锌离子的结合来形成一种“手指”状的结构域。

指状结构（finger structure） Michael Pique 和 Peter Wright 用核磁共振波谱确定非洲爪蟾的 Xfin 蛋白（某种 II 类启动子的激活因子）的锌指结构。他们发现图 12.1 中描述的并不像指形，或者说只是“一根粗短的手指”。他们还发现许多不同的指形蛋白具有相同构型却结合不同的特异 DNA 靶序列，因此认为此类指形结构自身并不能决定 DNA 结合特异性。这样只能是该指形结构或相邻区域的精确氨基酸序列决定了 DNA 结合序列的特异性。Xfin 锌指结构的一个 α 螺旋（图 12.1 的左侧）包含几个碱性氨基酸，它们看起来都位于与 DNA 接触的一侧。估计 α 螺旋结构中的这些氨基酸和其他氨基酸共同决定了该蛋白的 DNA 结合特异性。

Carl Pabo 及同事用 X 射线晶体衍射实验获得了 DNA 和小鼠蛋白 Zif268（TFIIIA 类锌指蛋白的一个成员）复合物的结构。小鼠蛋白 Zif268 被称为“立早蛋白”（immediate early protein），意即静止期细胞进入分裂期时被最早激活的基因之一。Zif268 蛋白有 3 个相邻的锌指结构，均嵌入 DNA 双螺旋的大沟中。在本章稍后我们将看到这 3 个锌指结构的

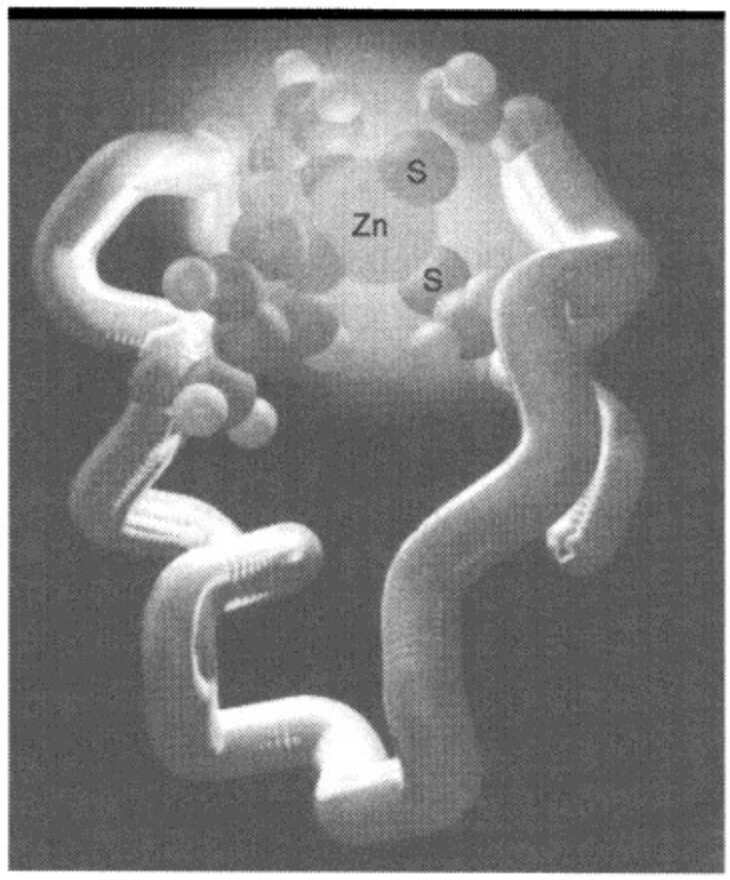


图 12.1　非洲爪蟾 Xfin 蛋白的一个锌指的三维立体结构。顶部中心的青绿色圆球代表锌，黄色圆球代表一对半胱氨酸的硫原子，左上角蓝色和绿色结构表示一对组氨酸。紫色管状结构表示锌指的骨架。（*Source*：Pique，Michael and Peter E. Wright，Dept. of Molecular Biology，Scripps Clinic Research Institute，La Jolla，CA. (cover photo，*Science* 245（11 Aug 1989).）

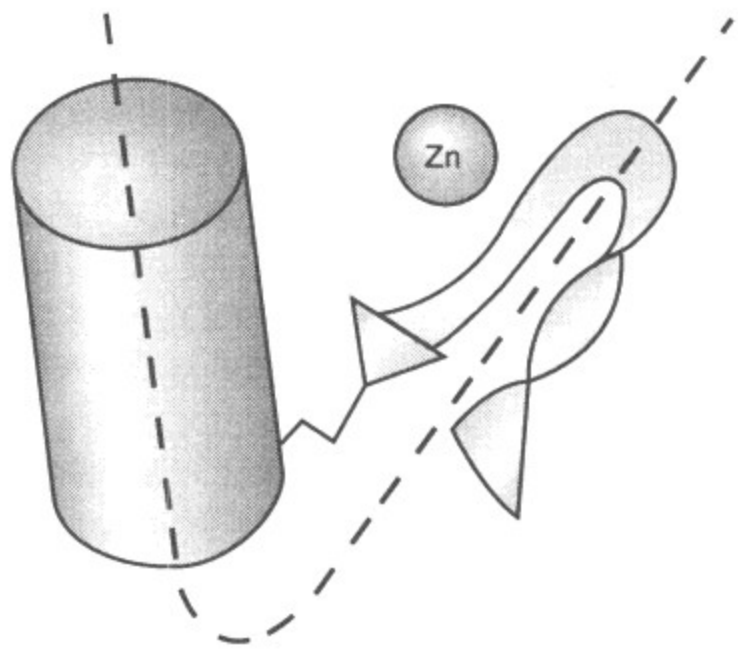


图 12.2　Zif268 蛋白锌指 1 的示意图。锌指结构右边是一个反向平行的β折叠（黄色），左边是α螺旋（红色）。以锌离子（蓝色）为中心连接了β折叠的一对半胱氨酸和α螺旋的一对组氨酸。虚线勾画出了“指形”轮廓。（*Source*：Adapted from Pavletich，N. P. and C. O. Pabo，Zinc finger-DNA recognition：Crystal structure of a Zif 268-DNA complex at 2.1 Å. *Science* 252：812，1991.）

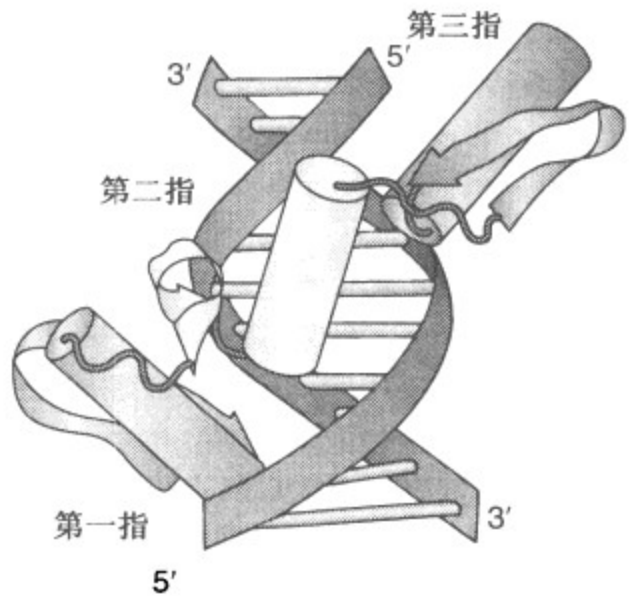


图 12.3　3 个 Zif268 锌指结构呈弯曲排列嵌入 DNA 大沟中。立体柱状结构代表α螺旋，带状结构表示β折叠。（*Source*：Adapted from Pavletich，N. P. and C. O. Pabo，Zinc finger-DNA recognition：Crystal structure of a Zif 268-DNA complex at 2.1 Å. *Science* 252：812，1991.）

排列。现在先来了解一下锌指本身的三维结构。图 12.2 展示了第一个锌指的结构。图中显示的指形构型并不明显，但通过近距离观察仍能沿着标记虚线看出大致的指形轮廓。与 Xfin 锌指相同，每个 Zif268 指结构左边有一个α螺旋，它通过底部的短环与锌指右边相连，形成反向平行的β折叠。不要混淆β折叠与指结构，它仅是指形结构的一部分。中间的锌离子（图中蓝色球）分别和α螺旋内部的一对组氨酸及β折叠中的一对半胱氨酸在空间形成对等构型。3 个指形结构几乎完全一致。

与 DNA 的相互作用　指形结构怎样与 DNA 靶序列相互作用呢？图 12.3 显示 Zif268 的 3 个锌指均与 DNA 双螺旋大沟接触。3 个指呈“C”形弯曲，与 DNA 双螺旋的凹槽匹配。所有指都以相同角度靠近 DNA，故蛋白质-DNA 接触的几何形状极为相似。每个指与 DNA 的结合依赖于α螺旋内氨基酸与 DNA 大沟碱基间的直接相互作用。详细的氨基酸-碱基相互作用见第 9 章。

与其他 DNA 结合蛋白的比较　对许多 DNA 结合蛋白的研究发现了一致的规律，即它们利用α螺旋与 DNA 大沟相互作用。Zif268 中的β折叠有什么作用呢？它可能与螺旋-转角-螺旋蛋白的第一个α螺旋的功能相同，即与 DNA 骨架相结合并帮助识别螺旋定位，从而利于与 DNA 大沟进行相互作用。

Zif268 与螺旋-转角-螺旋蛋白也有不同之处。后者的每个单体只有一个 DNA 结合域，而锌指蛋白的 DNA 结合域由一套组件构成，并由多个锌指与 DNA 相互作用。这意味着此类蛋白质无须像其

他 DNA 结合蛋白那样组成二聚体或四聚体后才能结合 DNA，它们自身有多个 DNA 结合域。此外，与螺旋-转角-螺旋蛋白一样，多数 DNA 结合蛋白只与 DNA 双螺旋的一条链而非两条链接触。而且，对于这种特定的指形蛋白而言，大多数接触发生在氨基酸和碱基之间，而不是与 DNA 的骨架之间。

1991 年，Nikola Pavletich 和 Carl Pabo 获得含 5 个锌指结构的人 GLI 蛋白与 DNA 的共晶体结构图。它与 3 个锌指结构的 Zif268 蛋白形成了有趣的对比，大沟依然是 DNA 和锌指接触的位点，但其中一个锌指（锌指 1）不与 DNA 接触。此外，这两个锌指-DNA 复合体的几何形状大体相似，锌指环绕 DNA 大沟，但在特定碱基和氨基酸间不存在简单的识别“代码”。

小结 锌指由一个反向平行的 β 折叠及紧邻的一个 α 螺旋组成。β 折叠中包含一对半胱氨酸，α 螺旋包含一对组氨酸，它们与锌离子螯合。氨基酸与金属离子的螯合有助于形成指状结构。指结构与其 DNA 靶序列的特异性识别发生于 DNA 双螺旋大沟内。

GAL4 蛋白

GAL4 蛋白是调节酵母半乳糖代谢基因的激活因子。GAL4 应答基因包括一个 GAL4 靶位点（转录起始位点上游的增强子区），这些靶位点被称为**上游激活序列**（upstream activating sequence，UAS_G）。GAL4 以二聚体形式结合在 UAS_G 上，其 DNA 结合模体位于蛋白质的前 40 个氨基酸中，二聚化模体位于第 50～94 位氨基酸残基之间。DNA 结合模体类似锌指，也包含锌离子和半胱氨酸残基，但其结构不同，表现在每个模体包含 6 个半胱氨酸但没有组氨酸，锌离子与半胱氨酸的比例是 1∶3。

Mark Ptashne、Stephen Harrison 及同事对 GAL4（只含前 65 个氨基酸）-DNA（人工合成的 17bp 寡聚脱氧核苷酸）复合体进行 X 射线晶体衍射实验，揭示了蛋白质-DNA 复合体的几个重要特性，包括 DNA 结合模体的形状、如何与靶 DNA 相互作用以及第 50～64 位氨基酸残基的部分二聚化模体。

DNA 结合模体 图 12.4 描述了二聚体 GAL4-DNA 复合体的结构。每个单体的一端包含一个 DNA 结合模体，该模体包含与两个锌离子（黄色球）结合的 6 个半胱氨酸，形成双金属巯基簇。每一模体均特征性地含有一个插入 DNA 双螺旋大沟的短 α 螺旋，并在该处进行特异性相互作用。每个单体的另一端是一个利于二聚化的 α 螺旋，这一点我们稍后进行讨论。

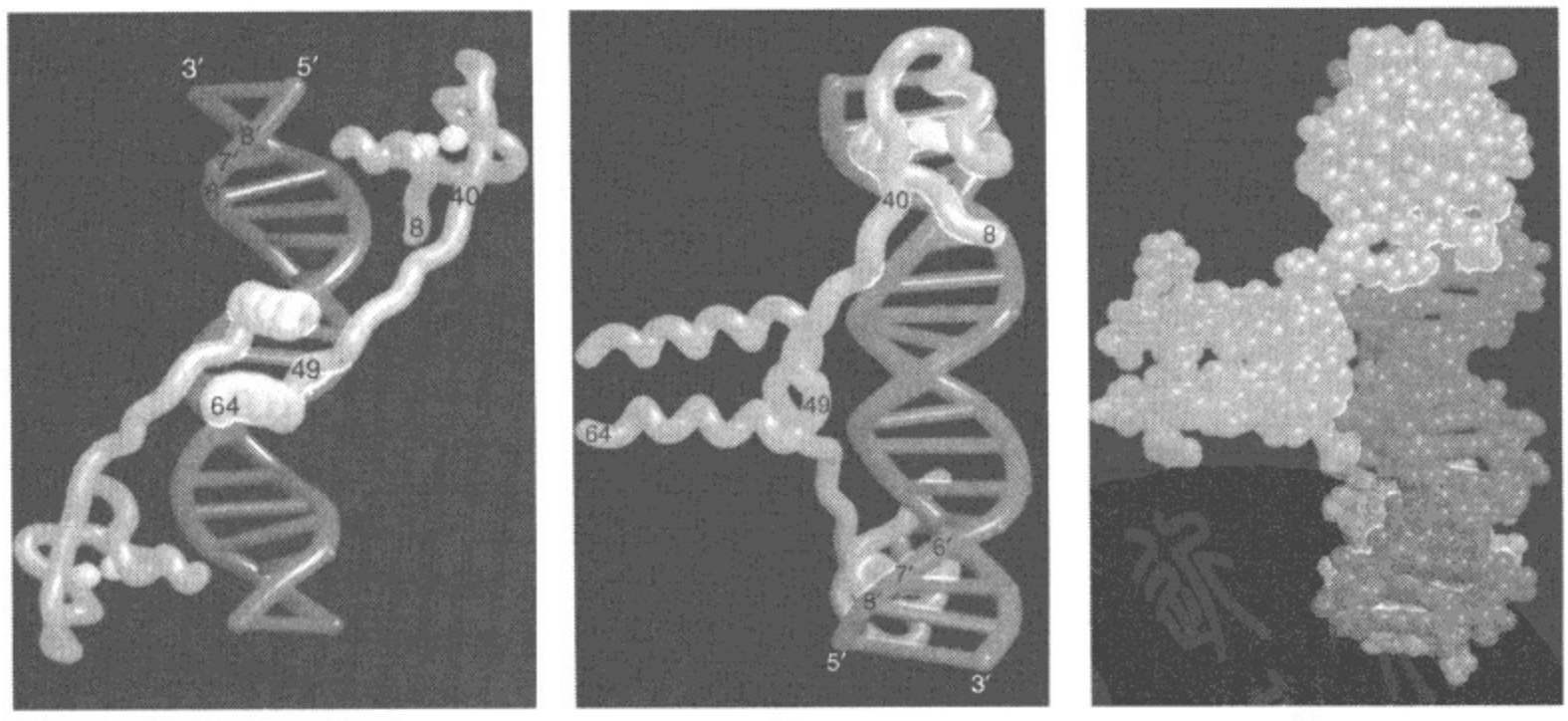


图 12.4 GAL4-DNA 复合体的三面观。（a）沿二重对称轴观察的复合体结构。红色代表 DNA，蓝色代表蛋白质，黄色球代表锌离子。3 个结构域的首尾氨基酸序号标记在上部的单体上。DNA 识别模体从第 8 位氨基酸延伸至第 40 位氨基酸，接头区在第 41～49 位氨基酸之间。二聚化域位于第 50～64 位氨基酸。（b）与（a）图垂直的角度观察的复合体结构。在左边中部，显示大体平行的二聚化元件。（c）与图（b）角度相同的空间填充模型。两个 GAL4 单体的识别组件分别与 DNA 链的正反面相接触。二聚化螺旋域和 DNA 小沟间有序契合。（*Source*：Marmorstein，R.，M. Carey，M. Ptashne，and S. C. Harrison，DNA recognition by GAL4：Structure of a protein-DNA complex. *Nature* 356（2 April 1992）p. 411，f. 3. Copyright © Macmillan Magazines Ltd.）

含锌的 DNA 结合模体如何发挥作用呢？图 12.5a 是这种模体的近观图，显示 6 个半胱氨酸、2 个锌离子（黄色球）和短识别螺旋。图 12.5b 详示了识别组件中的氨基酸和 DNA 碱基的相互作用。

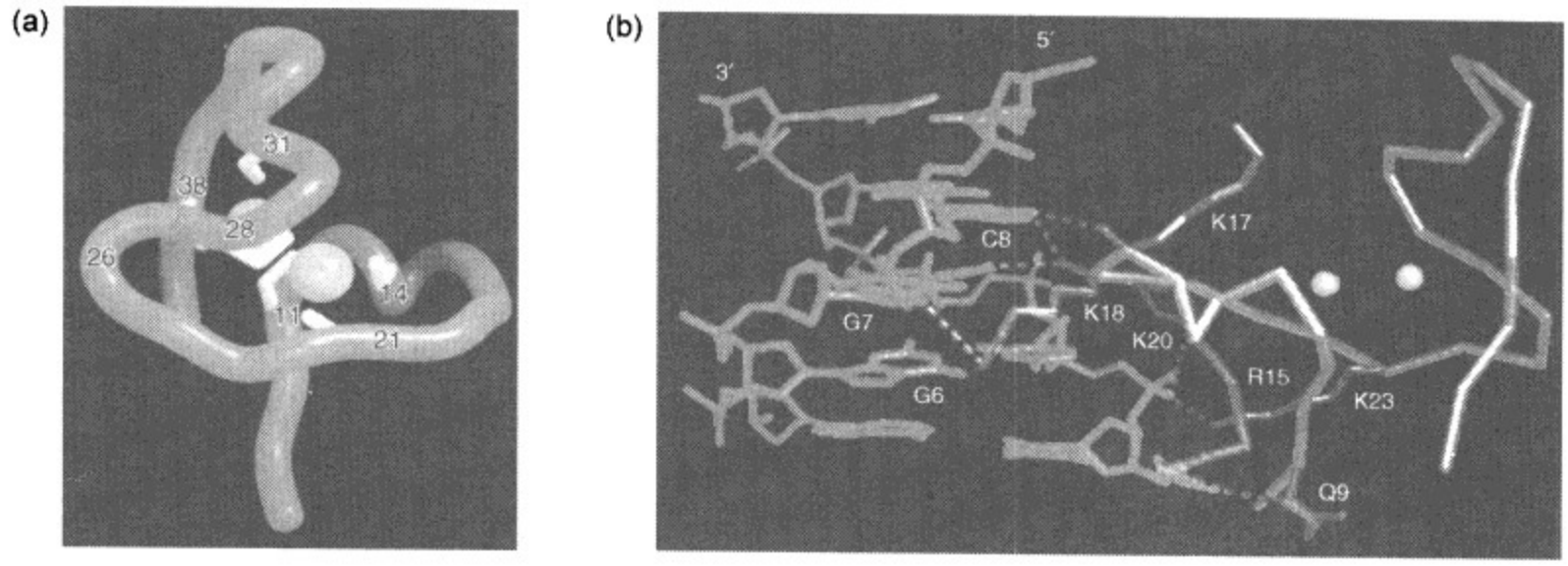


图 12.5　GAL4 蛋白的 DNA 识别组件。（a）识别组件模型。蓝色代表蛋白质骨架，黄色表示锌离子，白色表示 6 个半胱氨酸侧链（第 11、14、21、28、31 和 38）与锌离子配位结合，识别螺旋的短片段在右边，包括第 14 个氨基酸残基。（b）DNA-蛋白质的特异结合涉及 DNA 识别组件。蓝色代表蛋白质骨架，黄色代表锌离子。绿色表示与 DNA 碱基特异结合的氨基酸侧链，并标以氨基酸序号。图中显示了第 17 位精氨酸的侧链，如果 DNA 片段比图中显示的寡聚体更长一些，则该侧链就能与 DNA 上的磷酸基团形成氢键（绿色虚线表示）。[*Source*: Marmorstein, R., M. Carey, M. Ptashne, and S. C. Harrison, DNA recognition by GAL4: Structure of a protein-DNA complex. *Nature* 356 (2 April 1992) p. 412, f. 4-5. Copyright © Macmillan Magazines Ltd.]

二聚化模体　GAL4 单体利用 α 螺旋的二聚化作用在左侧形成平行的螺旋圈（coiled coil）（图 12.4b 和 c）。此图同时显示二聚化的 α 螺旋直指 DNA 小沟。在图 12.4 中，每个单体 DNA 识别组件和二聚化组件由一个伸展的区域相连。在本章后面讨论亮氨酸拉链结构和螺旋-环-螺旋模体时还会涉及螺旋二聚化模体的其他实例。

小结　GAL4 蛋白是 DNA 结合蛋白含锌家族的成员之一，但不具锌指结构。每个 GAL4 单体包含 6 个半胱氨酸及 2 个锌离子形成双金属巯基簇。DNA 识别组件的短 α 螺旋插入 DNA 双螺旋大沟并在那里进行特异性相互作用。GAL4 单体包含一个 α 螺旋二聚化模体，能在与其他 GAL4 单体相互作用时形成平行的螺旋卷曲。

细胞核受体

第三类含锌组件存在于**细胞核受体**（nuclear receptor）中。这些蛋白质与跨膜扩散的内分泌信号分子（类固醇和其他激素分子）相互作用，形成激素-受体复合物并结合到增强子或**激素响应元件**（hormone response element）上，激活相关基因的转录。与我们以前所了解的激活因子不同的是，这类激活因子必须结合一个效应分子（激素分子）才能发挥激活作用。这意味着它们必定有一个重要区域——激素结合域，实验结果也的确如此。

此类激素有性激素（雄性激素和雌性激素）、孕酮（一种孕激素，常见避孕药的基本成分）、糖皮质激素（皮质醇）、维生素 D（调控钙的代谢）、甲状腺激素和视黄酸（调控发育过程中基因的表达）。这些激素分别与相应受体结合，激活特定的一组基因。

传统分类法将核受体分为三类：**I 型受体**（type I receptor）包括类固醇激素受体，以**糖皮质激素受体**（glucocorticoid receptor）为代表。缺乏激素配体时，受体蛋白和其他蛋白偶联共存于胞质中。当激素配体与 I 型受体蛋白结合时，释放偶联蛋白，以同源二聚体形式进入细胞核，并结合到激素应答元件上。如糖皮质激素受体通常与热激蛋白 90（hsp90）偶联存在于胞质中，当它与激素配体结合后（图 12.6），构象发生改变，并与热激蛋白解离，随后进入细胞核，激活那些受增强子［又称**糖皮质激素应答元件**（glucocorticoid response element，**GRE**）］所调控的基因。

Sigler 及同事对糖皮质激素受体分别结合两种寡核苷酸的共结晶复合体进行了 X 射线衍射实验。其中一个寡核苷酸含有 2 个半-靶位点（two target half-site），中间被 3 个碱基隔开（正常情况），另一个寡核苷酸与前者不同的是 2 个半-靶位点被 4 个碱基隔开（多一个额外碱基）。正是这一个碱基的差别，导致 DNA 与蛋白质相互作用的本质产生了很大差异，稍后我们会谈及。

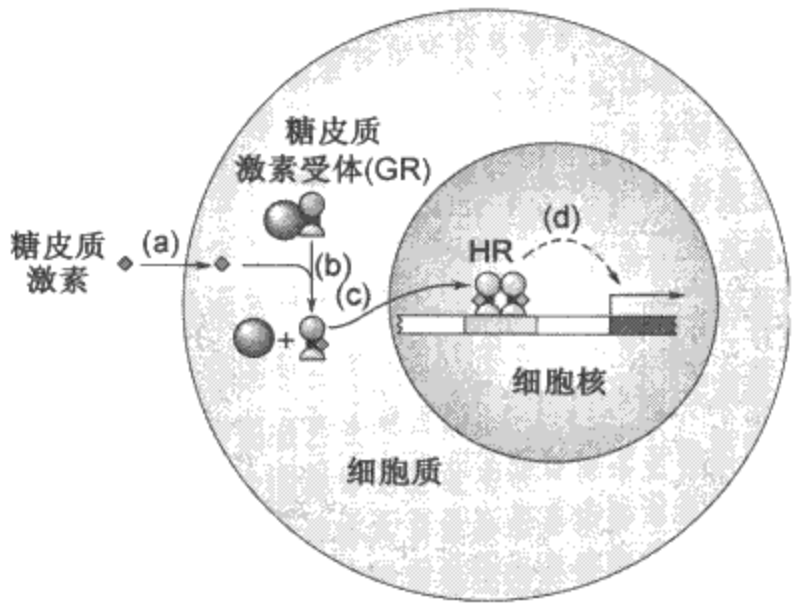


图 12.6　糖皮质激素的作用。糖皮质激素受体（GR）与热激蛋白90（hsp90）结合，以无活性形式存在于细胞质中。(a) 糖皮质激素（蓝色）通过细胞膜扩散进入胞质；(b) 糖皮质激素与受体（GR）结合（分别用红色和绿色表示），构象发生变化，释放 hsp90（橘黄色）；(c) 激素-受体复合物（HR）进入细胞核，与另一个 HR 结合成二聚体，然后与位于激素激活基因（棕色）上游的激素响应元件或增强子（粉色）结合；(d) HR 二聚体与增强子的结合，激活（虚线箭头）相关基因的转录（弯箭头）。

图 12.7a 是蛋白质与 DNA 相互作用的立体结构，显示 2 个半-靶位点间含有一个额外碱基对的寡核苷酸与蛋白质间的相互作用。该复合体有几个明显的特点：①结合域二聚化，其中一个单体（红色）与 DNA 形成特异性结合，而另一个单体（蓝色）与 DNA 形成非特异结合。在天然 DNA 双螺旋中，靶序列半结合位点之间通常被 3 个碱基隔开，使两个蛋白质单体均与 DNA 之间形成专一性结合。但一对额外碱基阻碍了这种完美结合。②每个结合模体都包含 2 个锌离子（黄色），不像典型的锌指结构那样只含 1 个锌离子。③每个锌离子与 4 个半胱氨酸形成类指（finger-like）（图中不明显）。④结合域中的氨基端指参与大部分与靶序列的相互作用。这类相互作用总涉及 α 螺旋，它们在图上看起来像一串挤在一起的方框结构。这种 α 螺旋处于特异性结构域（红色）上部锌离子的左边，也正好位于非特异性结构域（蓝色）下部锌离子的左边。图 12.7b 显示了识别螺旋和 DNA 靶序列间特异性氨基酸和碱基间的结合。在螺旋区外的一些氨基酸也通过 DNA 骨架中的磷酸与 DNA 相结合。

II 型受体　**甲状腺激素受体**（thyroid hormone receptor）是典型的 II 型受体，存在于细胞核中，与**视黄酸受体 X**（retinoic acid receptor X，**RXR**）组成二聚体，而其中的 X 以 9 个顺式视黄酸（9-*cis* retinoic acid）为配体。这种二聚体在有或无激素配体时，都能结合靶序列。在第 13 章我们将看到，缺乏配体时，靶序列与 II 型受体的结合会抑制转录，而结合有配体的 II 型受体则会刺激转录。因此，环境条件决定了相同的蛋白质成为激活因子或抑制因子的可能性。

III 型受体　目前对 III 型受体研究得还不十分清楚。由于其配体有待鉴定，因此也被称为“孤儿受体”（orphan receptor）。对此类受体的深入研究也许会将其部分或全部归属到 I 型或 II 型受体中。

最后需要注意的是，3 种含锌 DNA 结合组件几乎都利用一个共同的模体与 DNA 靶位点相互作用，这就是 α 螺旋。

小结　I 型核受体与另一种蛋白质结合并位于细胞质中。当外源激素配体与 I 型受体结合后，释放偶联的胞质蛋白，并进入核内与增强子结合，行使激活因子的功能。糖皮质激素受体是该类群的代表，它的 DNA 结合域有 2 个含锌组件。一个组件含有大部分与 DNA 结合相关的氨基酸（在 α 识别螺旋中），另一组件则提供了供蛋白质-蛋白质相互作用形成二聚体的触点。II 型核受体位于细胞核，结合在其 DNA 靶位点上。在缺乏配体时，它们抑制基因活性；与配体结合后，则激活转录。III 型受体是“孤儿”受体，它们的配体还不清楚。

同源结构域

同源结构域（homeodomain）是在一类激活因子大家族中发现的 DNA 结合域。由于其编码基因的区域为**同源盒**（homeobox）而得名。同源盒最早发现于果蝇的被称为同源基因的调控基因中，该基因的突变会引起果蝇肢体的移位畸形。例如，被称为触角足（antennapedia）的突变体，腿长在原来触角所在的位置（图 12.8）。

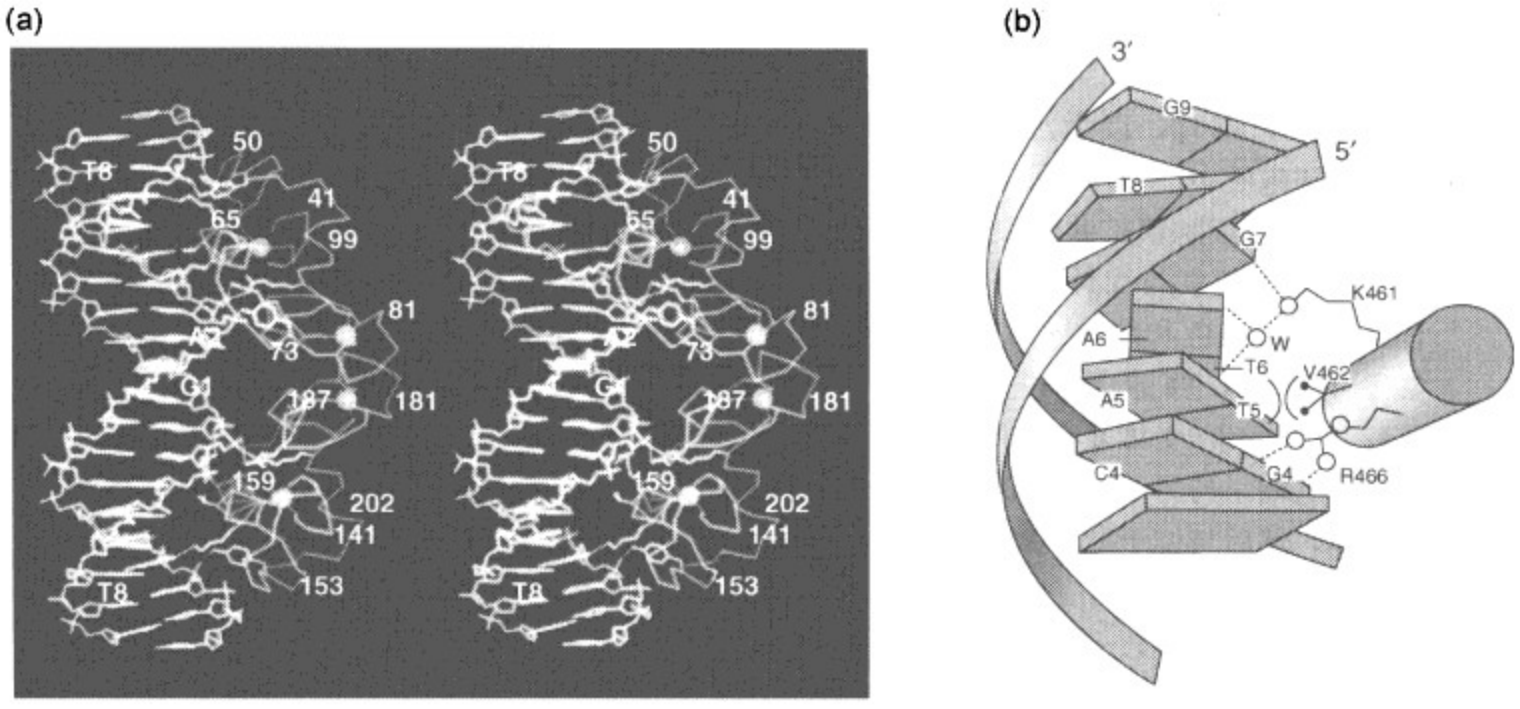


图 12.7　糖皮质激素受体 DNA-结合域与其靶 DNA 间的结合。（a）显示氨基酸/DNA 以及蛋白质-蛋白质相互作用的立体图。左边是 DNA，右边是蛋白质的 β-碳链。由于在 2 个靶序列半结合位点之间含有一对多余碱基，使蛋白质二聚体的一个单体（红色）与 DNA 特异结合，另一个单体（蓝色）与 DNA 非特异结合。黄色球代表锌离子，深绿色表示半胱氨酸侧链，浅绿色表示氨基酸与 DNA 的相互作用。用带数字的字母和数字分别标记 DNA 碱基和氨基酸。识别螺旋位于最上部和最下部的锌离子的左边。注意，二聚化域的相互作用发生在中间两个锌离子之间。（b）氨基酸和碱基之间的特异识别涉及识别螺旋。一个水分子（W）介导 461 位赖氨酸与 DNA 的一些氢键结合。（*Source*：Luisi，B. F.，W. X. Xu，Z. Otwinowski，L. P. Freedman，K. R. Yamamoto，and P. B. Sigler，Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* 352（8 Aug 1991）p. 500，f. 4a. Copyright © Macmillan Magazines Ltd.）

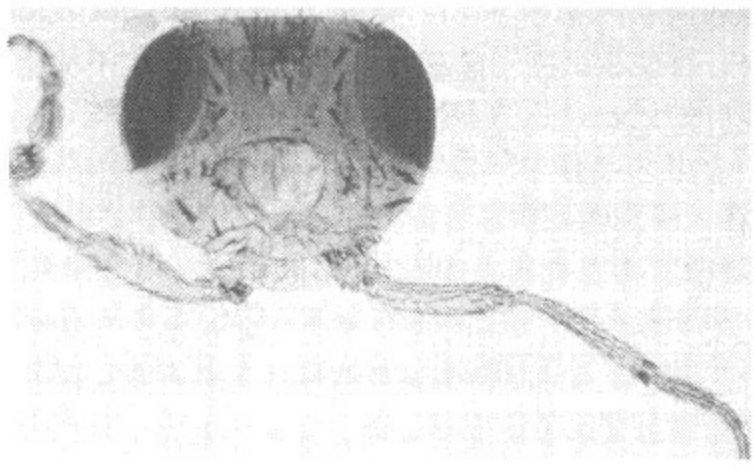


图 12.8　果蝇触角足表型。头部原来生长触角的位置长出了足。（*Source*：Courtesy Walter J. Gehring，University of Basel，Switzerland.）

同源结构域蛋白是 DNA 结合蛋白中的螺旋-转角-螺旋家族成员。每个同源结构域蛋白包括 3 个 α 螺旋，第二、三螺旋形成螺旋-转角-螺旋模体，第三个螺旋作为识别螺旋起作用。但大多数同源结构域蛋白的 N 端还有一个不同于螺旋-转角-螺旋的臂，可插入 DNA 小沟。图 12.9 显示来自果蝇的一个典型同源框（由同源框基因 *engrailed* 编码）与 DNA 靶序列间的相互作用。此图来自于 Thomas Kornberg 和 Carl Pabo 对 engrailed 蛋白的同源结构域，与含有其对应结合位点的寡核苷酸复合物共结晶的 X 射线衍射分析结果。同源结构域蛋白与 DNA 结合的特异性较弱，需要其他蛋白质协助才能高效、专一地结合目标序列。

小结　真核细胞转录激活因子的同源结构域包含 DNA 结合模体，其作用原理类似原核细胞的螺旋-转角-螺旋模体，由一个识别螺旋嵌入 DNA 大沟并进行特异相互作用，其 N 端臂则插入相邻的 DNA 小沟内。

亮氨酸拉链和螺旋-环-螺旋域

如前述 DNA 结合域一样，亮氨酸拉链和螺旋-环-螺旋域（bZIP and bHLH domain）有两个功能：结合 DNA 和二聚化作用。其中 ZIP 和 HLH 分别指域中的**亮氨酸拉链**（leucine zipper）和**螺旋-环-螺旋**（helix-loop-helix）部分，均为二聚化模体（dimerization motif）。“b”指每个域中的碱性区域，构成 DNA 结合模体的主体。

我们首先以亮氨酸拉链为例来了解这种兼具二聚化和结合 DNA 功能结构域。该域由两个多肽链组成：α 螺旋中每隔 7 个氨基酸就出现一个亮氨酸

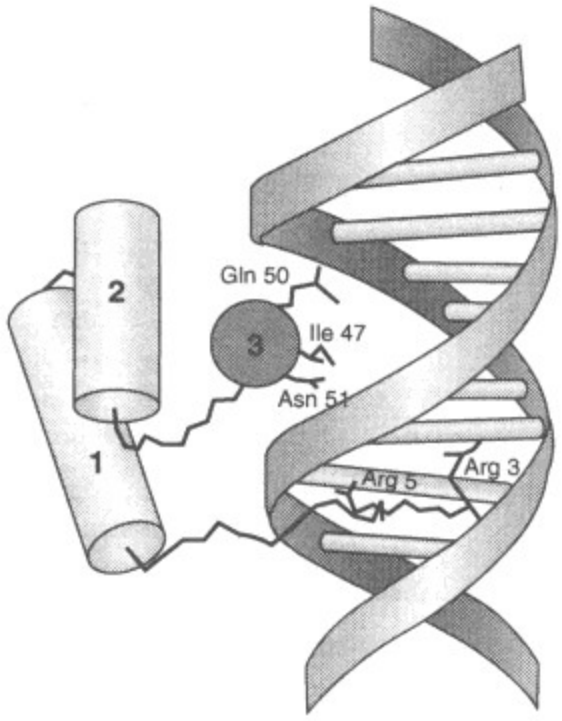


图 12.9　同源结构域-DNA 复合体结构图。标有数字的 3 个螺旋位于左边，DNA 靶序列位于右边。显示识别螺旋的端部（3，红色），位于 DNA 大沟内。N 端长臂插入 DNA 小沟中，显示关键氨基酸侧链与 DNA 相互作用。（*Source*：Adapted from Kissinger，C. R.，B. Liu，E. Martin-Blanco，T. B. Kornberg，and C. O. Pabo，Crystal structure of an engrailed homeodomain-DNA complex at 2. 8 Å resolution：A framework for understanding homeodomain-DNA interactions. *Cell* 63（2 November，1990）p. 582. f. 5b.）

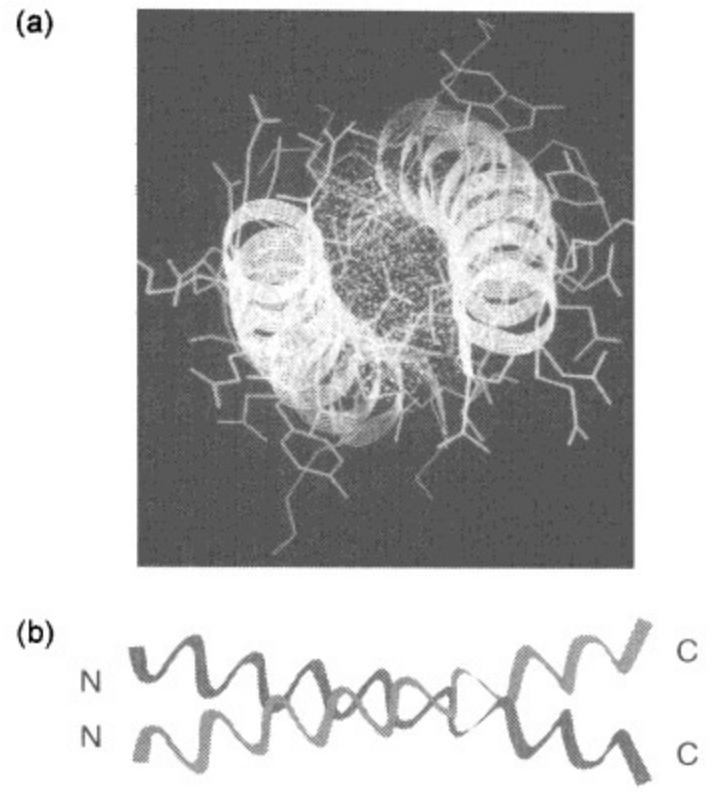


图 12.10　亮氨酸拉链结构。（a）Kim 和 Alber 及同事对一个含亮氨酸拉链模体的转录激活因子 GCN4 中的 33 个氨基酸肽段进行了 X 射线晶体衍射实验。晶体沿拉链二聚体螺旋指向纸面外。（b）两个 α 螺旋（红色和蓝色）组成二聚体的侧面图。因为两个 α 螺旋的 N 端都在左边，因此组成一个平行的缠绕螺旋。（*Source*：（a）O'Shea，E. K.，J. D. Klemm，P. S. Kim，and T. Alber，X-ray structure of the GCN4 leucine zipper，a two-stranded，parallel coiled coil. *Science* 254（25 Oct 1991）p. 541，f. 3 Copyright © AAAS.）

残基（或其他疏水性氨基酸），因此这些氨基酸残基都位于螺旋的同一侧。这种排列非常有利于两个相同蛋白质单体间的相互作用，使两个 α 螺旋成为拉链的两边。

为搞清亮氨酸拉链的结构，Peter Kim 和 Tom Alber 及同事按照 GCN4（酵母中调节氨基酸代谢的转录激活因子）的 bZIP 域合成了一种人工多肽，并获得了共晶体结构。X 射线衍射实验表明，亮氨酸拉链二聚化域像一个螺旋线圈（图 12.10）。由于氨基端至羧基端的方向一致（b 图，从左至右）使两个 α 螺旋形成平行结构。图 12.10a 中的螺旋线圈直接伸出纸平面指向读者，使螺旋线圈中的超螺旋程度清晰地呈现出来。注意它与 GAL4 螺旋二聚体模体的相似之处（图 12.4）。

以上晶体图主要显示未结合 DNA 的拉链结构，没有提供蛋白质与 DNA 结合的任何信息。Kevin Struhl 和 Stephen Harrison 及同事对结合在 DNA 靶序列上的 GCN4 激活因子中的亮氨酸拉链域所做的 X 射线晶体衍射实验弥补了这一缺陷。图 12.11 表明，亮氨酸拉链不仅将两个单体结合在一起，还将结构域的两个碱性区置于合适位置，恰似镊子或钳子将 DNA 紧紧抓住，其中的碱性模体嵌入 DNA 大沟内。

Harold Weintraub 和 Carl Pabo 及同事获得了激活因子 MyoD 的 bHLH 域与其 DNA 靶位点结合的复合体晶体结构，与我们前面讨论的亮氨酸拉链-DNA 复合体结构极其相似（图 12.12）。螺旋-环-螺旋是二聚化模体，但每个螺旋-环-螺旋域的长螺旋（螺旋 1）包含结构域的碱性区域，像亮氨酸拉链那样，碱性区与 DNA 大沟结合。

有些蛋白质，如癌基因产物 Myc、Max 有 **bHLH-ZIP** 域，其 HLH 和 ZIP 模体都邻近一个碱性模体。bHLH-ZIP 域与 DNA 的作用方式类似 bHLH 与 DNA 的作用方式。主要差别在于 bHLH-ZIP 域可能需要额外亮氨酸拉链的作用，以确保蛋白质单体的二聚化。

小结　bZIP 蛋白通过亮氨酸拉链而二聚化，使每个单体的相邻碱性区域像一把钳子一样夹住 DNA 靶序列。类似地，bHLH 蛋白通过螺旋-环-螺旋模体而二聚化，使每个长螺旋的碱性部分

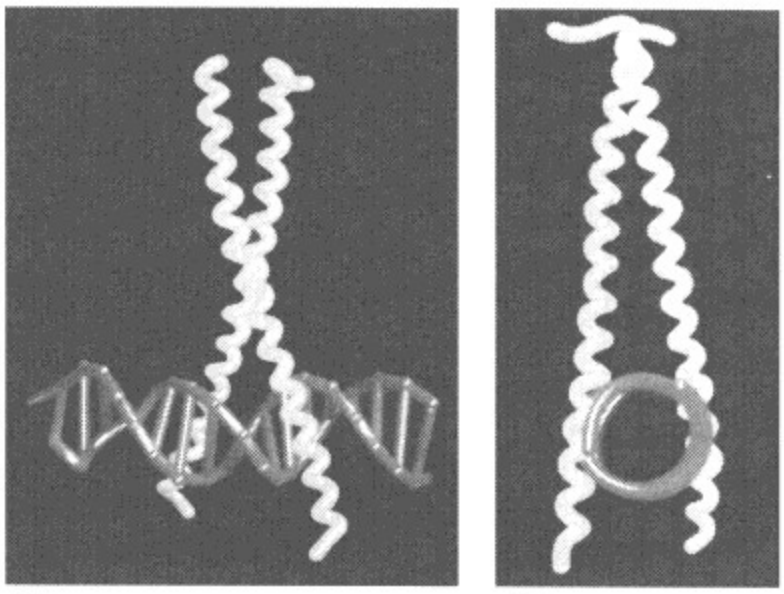


图 12.11　GCN4-DNA 复合体的亮氨酸拉链模体的晶体结构。DNA（红色）含亮氨酸拉链模体（黄色）识别的靶序列。注意蛋白质单体之间相互作用的螺线圈的本质以及紧抓 DNA 的蛋白质分子的钳状外形。(a)DNA 的侧面观；(b)DNA 的端面观。（*Source*：Ellenberger，T. E.，C. J. Brandl，K. Struhl，and S. C. Harrison，The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices：Crystal structure of the protein-DNA complex. *Cell* 71（24 Dec 1992）p. 1227，f. 3a-b. Reprinted by permission of Elsevier Science.）

抓住 DNA 靶序列，就像 bZIP 蛋白那样。BHLH 和 bHLH-ZIP 域都以相同方式结合 DNA 靶序列，但后者由于其亮氨酸拉链而有额外的二聚化潜力。

12.3　激活因子各结构域的独立性

我们已经看到了几种激活因子的 DNA 结合域和转录激活域。蛋白质的这些结构域彼此独立折叠，形成特定的三维结构，独立行使功能。为了揭示其独立性，Roger Brent 和 Mark Ptashne 利用一种蛋白质的 DNA 结合域和另一种蛋白质的转录激活域构建了一种嵌合体（chimeric）。该杂合蛋白质仍能作为激活因子起作用，其特异性由 DNA 结合域决定。

他们用编码 GAL4 和 LexA 两种蛋白质的基因构建了重组体。我们已经学习了 GAL4 的 DNA 结合域和转录激活域。LexA 是原核生物的阻遏物，与 *lexA* 操纵子结合并阻抑下游基因的转录。LexA 没有转录激活域，没有转录激活功能。他们构建了包含 GAL4 转录激活域和 LexA DNA 结合域的重组

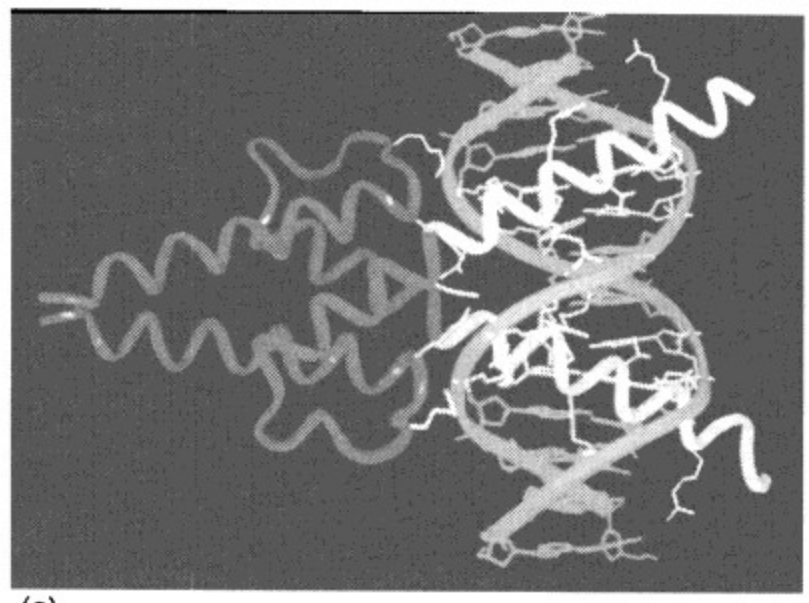


图 12.12　MyoD 的 bHLH 域和靶 DNA 序列复合体的晶体结构。(a) α 螺旋以卷曲带状结构表示；(b) α 螺旋以圆柱体表示。（*Source*：Ma，P. C. M.，M. A. Rould，H. Weintraub，and C. O. Paol，Crystal structure of MyoD bHLH domain-DNA complex：Perspectives on DNA recognition and implications for transcriptional activation. *Cell* 77（6 May 1994）p. 453，f. 2a. Reprinted by permission of Elsevies Science.）

基因。为了分析重组基因所编码蛋白质产物的活性，他们将如下两个质粒转入酵母细胞。一个包含编码 LexA-GAL4 融合蛋白的重组基因，产生杂合产物。另一个包含能响应 GAL4 的启动子（*GAL1* 基因或 *CYC1* 基因启动子），并连接有大肠杆菌 β-半乳糖苷酶作为报告基因（第 5 章）。GAL4 响应启动子的转录越活跃，产生的 β-半乳糖苷酶越多。通过测定 β-半乳糖苷酶的量可检测目的基因的转录效率。

完成这一实验还需要嵌合蛋白的 DNA 结合位点。GAL4 通常可以结合到 UAS_G 上游的增强子序列。但嵌合蛋白不能识别这个位点，该蛋白质只有 LexA DNA 结合域。为了诱导 *GAL1* 启动子对激活的响应，需要引入 LexA 基因 DNA 结合域的靶序

列。因此，他们用 *lexA* 操纵子替代了 UAS_G 序列。值得注意的是，酵母细胞中没有 *lexA* 操纵子，只是为了实验的目的才进行了以上操作。现在的问题是：嵌合蛋白能激活 *GAL1* 基因转录吗？

答案是肯定的，图 12.13 对此做了说明。3 个用于测试的质粒分别包含 UAS_G、无靶结合位点、含 *lexA* 操纵子。转录激活因子是 LexA-GAL4（如前所述），或 LexA（作为阴性对照）。因为酵母细胞自身可产生 GAL4，它可通过 UAS_G 行使转录激活作用，因此在 UAS_G 存在时（图 12.13a），无论用哪种激活因子，都可以产生大量 β-半乳糖苷酶。而缺乏靶 DNA 结合位点时，酵母细胞不产生 β-半乳糖苷酶（图 12.13b）。最后，*lexA* 操纵子替换 UAS_G 序列（12.13c），LexA-GAL4 嵌合蛋白激活的 β-半乳糖苷酶的表达效率是 LexA 的 500 倍。因此，用一种与 GAL4 完全不相关的蛋白质结合域取代 GAL4 的结合域，可产生有活性的转录激活因子。这表明 GAL4 的转录激活域和 DNA 结合域可以彼此独立行使功能。

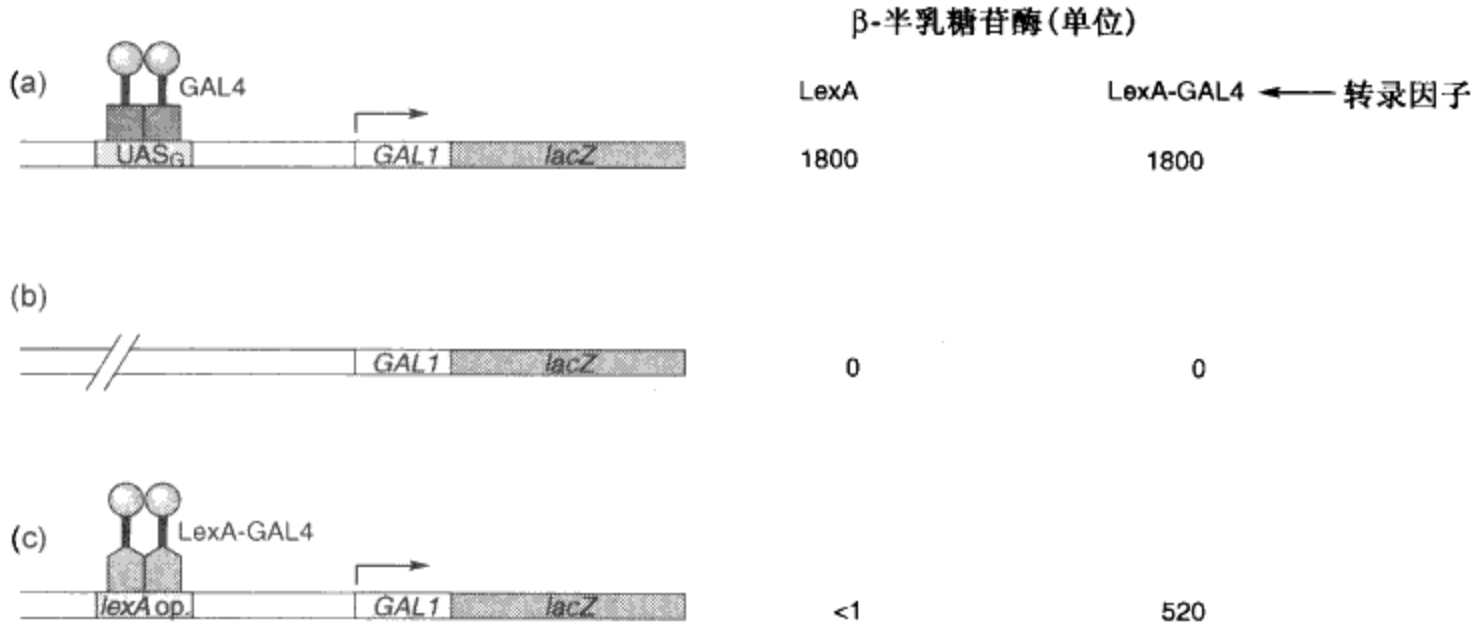


图 12.13　嵌合激活因子的转录活性。Brent 和 Ptashne 将两种质粒导入酵母细胞。①编码 LexA-GAL4 蛋白的质粒，产生包含 GAL4 转录激活域（绿色）和 LexA 的 DNA 结合域（蓝色）的杂合蛋白；②图 a～c 显示 3 种质粒，每种质粒的 *GAL1* 基因启动子都与大肠杆菌 *lacZ* 报告基因相连。以 β-半乳糖苷酶（右侧给出）的产量指示启动子活性。（a）有 UAS_G 元件时，转录非常活跃，且不依赖新添加的转录激活因子，因为内在的 GAL4 可通过 UAS_G 激活转录；（b）没有 DNA 靶结合位点时，LexA-GAL4 融合蛋白不能与 *GAL1* 启动子附近的序列结合，不激活转录；（c）*lexA* 操纵基因存在时，LexA-GAL4 融合蛋白可以极大地激活基因转录。LexA 的 DNA 结合域可与 *lexA* 操纵基因结合，而 GAL4 转录激活域可通过 *GAL1* 启动子增强目的基因的转录。

小结　激活因子蛋白的 DNA 结合域和转录激活域是相互独立的组件。可以构建一种杂合蛋白，兼具一种蛋白质的 DNA 结合域和另一种蛋白质的转录激活域，并证明此杂合蛋白仍具有激活因子的功能。

12.4　激活因子的功能

在原核生物中，RNA 聚合酶核心酶无法起始实质性的转录，而 RNA 聚合酶全酶能催化本底水平的转录。但弱启动子引发的本底水平转录常常不能满足正常的生理需求，因此细胞中的转录激活因子通过所谓的召集作用（recruitment）来提高转录水平。召集作用能促使 RNA 聚合酶全酶与启动子的紧密结合。

真核生物转录激活因子也介导 RNA 聚合酶全酶与启动子结合，但不像原核生物转录激活因子那样直接。真核转录激活因子刺激通用转录因子及 RNA 聚合酶与启动子结合。图 12.14 中的两个假说或许可以解释这种召集作用：①通用转录因子促使前起始复合物逐步形成；②通用转录因子和其他蛋白质与 **RNA 聚合酶 II 全酶**结合后，一起被召集到启动子上。实际情况也许是两种假说的结合。无论如何，召集作用的发生需要通用转录因子和转录激活因子的直接接触［尽管在本章后面，我们会看到一些激活因子需要借助辅激活因子（coativator）的帮助与通用转录因子结合］。转录激活因子将与哪些通用转录因子接触呢？研究结果显示很多因子都可能是其结合的目标，但最早发现的是 TFIID 因子。

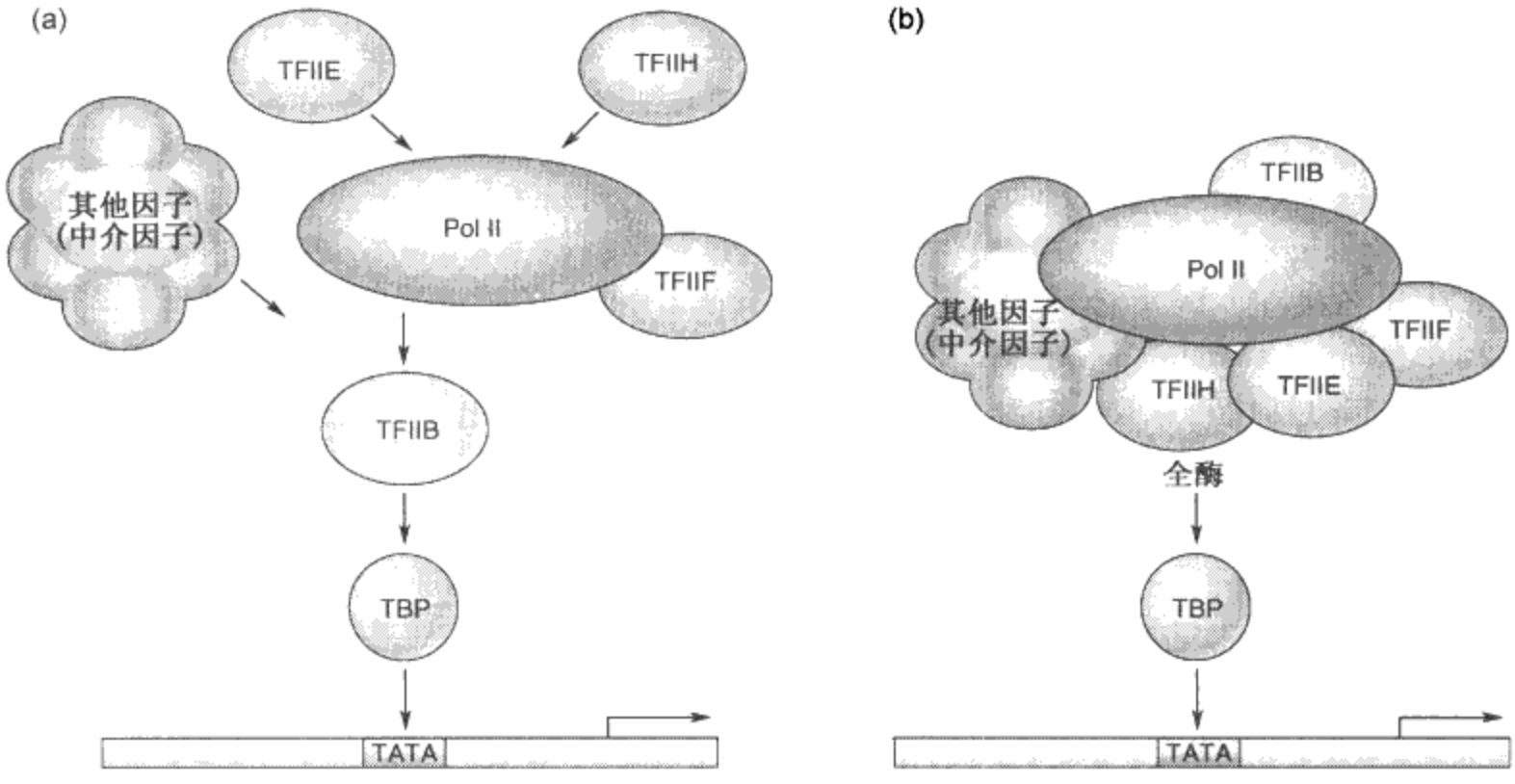


图 12.14　酵母前起始复合物组分的两种召集模型。（a）召集作用的传统观点。在体外，前起始复合物组分的逐步加入。CTD 指聚合酶 II 最大亚基的 C 端结构域。（b）聚合酶全酶的召集。TBP 首先与靶序列结合，然后聚合酶全酶结合，形成前起始复合体。（*Source*：Adapted from Koleske，A. J. and R. A. Young，An RNA polymerase II holoenzyme responsive to activators. *Nature* 368：466，1994.）

TFIID 的召集作用

Keith Stringer、James Ingles 和 Jack Greenblatt 在 1990 年用一系列实验鉴定了能与疱疹病毒转录因子 **VP16** 的酸性转录激活域结合的转录因子。研究人员表达了 VP16 转录激活域和金黄色葡萄球菌蛋白 A 的融合蛋白，其中金黄色葡萄菌蛋白 A 可以专一地与免疫球蛋白 IgG 紧密结合。将这种融合蛋白（或仅含蛋白 A）固定在 IgG 凝胶柱上，并用这种亲和层析柱筛选与 VP16 转录激活域相互作用的蛋白质。实验中他们将 HeLa 细胞核提取物通过蛋白 A 层析柱或蛋白 A/VP16 转录激活域融合蛋白凝胶层析柱，然后用截断转录（第 5 章）验证各种洗脱组分在体外精确转录腺病毒主要晚期基因的能力。结果显示，流过蛋白 A 柱的提取液仍足以支持转录，表明没有重要因子与蛋白质 A 非特异性结合。但是，流过蛋白 A/VP16 激活域蛋白柱的提取液无转录激活能力，除非将结合于亲和柱中的蛋白质重新加入。结果表明，某种或某些体外转录必需因子与 VP16 转录激活域发生了结合。

Stringer 及同事在早先的实验中发现 TFIID 是体外转录体系的限速因子。由此他们猜测与亲和柱结合的蛋白质是 TFIID。为了验证这一点，他们用加热的方法耗尽核提取物的TFIID，然后将结合蛋白 A 亲和柱或蛋白 A/VP16 转录激活域的亲和柱的成分加回到热处理过的核提取物中。图 12.15 表明，蛋白 A 柱的结合物不能恢复 TFIID 耗尽的核提取物的转录活性，而蛋白 A/VP16 转录激活域亲和柱的结合物可恢复其转录活性。说明 TFIID 是结合 VP16 转录激活域的因子。

为进一步验证上述结论，Stringer 及同事首先证明结合 VP16 转录激活域的物质在 DEAE 纤维素离子交换层析柱的表现与 TFIID 因子相似。然后，用 VP16 转录激活域亲和柱的结合物测试在模板转换实验（template commitment experiment）中取代 TFIID 的能力。先在一个模板上形成转录前起始复合物，然后加入另一个模板看是否能被转录。在该实验条件下，转录第二个模板需依赖 TFIID 因子。结果发现，VP16 转录激活域结合物可以将转录转换到第二个模板上，而蛋白 A 柱的结合物则不具有这种功能。利用酵母细胞核提取物的相同实验也提供了强有力的证据，说明该实验体系中 TFIID 是 VP16 转录激活域的重要靶因子。

小结　在亲和层析条件下，疱疹病毒的转录激活因子 VP16 的酸性激活域与 TFIID 结合。

聚合酶全酶的召集

在第 11 章我们学过，RNA 聚合酶 II 能以全酶形式从真核细胞中分离出来。全酶是包含一组通用转录因子和其他多肽的复合体。迄今为止，我们的讨论都基于转录激活因子以每次召集一个通用转录因子的方式形成起始复合物前体。但也许转录激活

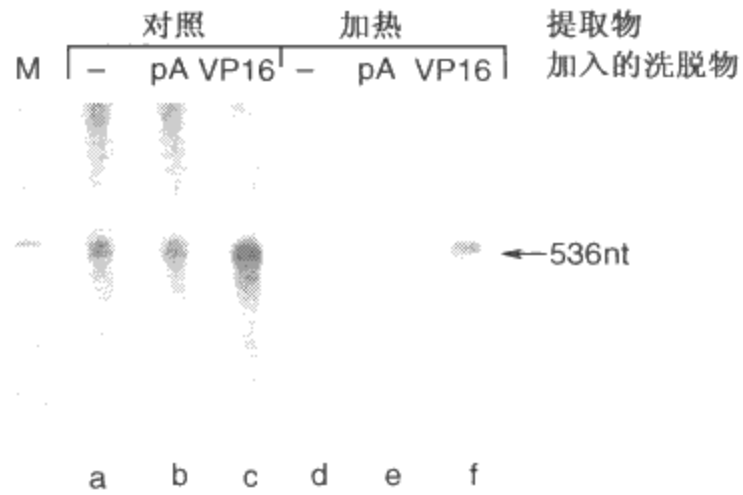


图 12.15　酸性转录激活域和 TFIID 结合的证据。Stringer 及同事用含有蛋白 A/VP16 转录激活域融合蛋白，或仅含蛋白 A 的亲和层析柱对 HeLa 细胞提取物进行分离。然后洗脱结合在层析柱上的蛋白质，测试洗脱物对热处理破坏了 TFIID 的核提取物恢复体外失控转录的能力。泳道a～c 为核提取物未经热处理的对照。TFIID 有活性，所有泳道都显示转录活性。泳道 d～f 为加热处理的核提取物，并加入下列物质：不加任何因子（–）、来自蛋白 A 柱的洗脱物（pA）、来自融合蛋白柱的洗脱物（VP16）。只有来自 VP16 融合蛋白柱的洗脱物可替代缺失的 TFIID，精确起始失控转录，获得预期长度（536nt）的转录物。因此，在亲和层析柱中，TFIID 一定与 VP16 转录激活域结合了。（*Source*：Stringer，K. F.，C. J. Ingles，and J. Greenblatt，Direct and selective binding of an acidic transcriptional activation domain to the TATA-box factor TFIID. *Nature* 345 (1990) f. 2，p. 784. Copyright © Macmillan Magazines Ltd.）

因子将全酶作为一个整体来召集，其他的少数因子再单独结合到启动子上。有证据表明，的确存在对全酶的召集。

1994 年，Anthony Koleske 和 Richard Young 从酵母细胞中分离出含聚合酶 II、TFIIB、TFIIF 和 TFIIH，以及 SRB2、4、5 和 6 的全酶。他们进一步证实这种全酶在 TBP 和 TFIIE 协助下，可以在体外精确转录带有 *CYC1* 基因启动子的模板。还证明嵌合转录激活因子 GAL4-VP16 可以激活此全酶的转录。由于实验提供的是完整聚合酶全酶，因此，后面这一发现提示激活因子召集完整的聚合酶全酶到启动子上，而不是在启动子上一次一个地组装（图 12.14）。

1998 年，研究者已从多种不同生物中分离到聚合酶全酶，其蛋白质组成各异。有些全酶含有大多数或者全部通用转录因子和许多其他蛋白质。Koleske 和 Young 推测，酵母聚合酶全酶包含 RNA 聚合酶 II、中介物（mediator，辅激活因子复合物）和除了 TFIID、TFIIE 外的所有通用转录因子。理论上，此聚合酶全酶可作为前体单元整体被召集，或分批被召集。

聚合酶作为整体被召集的证据　1995 年，Mark Ptashne 及同事为聚合酶全酶的召集模型增添了另一个强有力的证据。他们做了如下分析，如果全酶以整体被召集，则激活因子（结合在启动子附近）的任意部分和全酶的任意部分之间的相互作用应该能召集全酶到启动子上。这种蛋白质间的相互作用无须激活因子的转录激活域，也无须激活因子在通用转录因子的靶位点参与，激活因子和聚合酶全酶之间的任何接触都应导致转录激活。而另一方面，如果前起始复合物一次一个蛋白质地进行组装，那么在激活因子和全酶的非关键成分间的非正常互作就不会激活转录。

Ptashne 及同事利用一次碰巧机会验证了这些预测。他们以前曾分离到一种酵母突变体，其全酶蛋白（GAL11）中的点突变改变了单个氨基酸。他们将该蛋白质命名为 GAL11P（P 表示增效子 potentiator）。因为它对激活因子 GAL4 的弱突变型产生了很强的响应。结合生化和遗传方法，他们找到了 GAL11P 增效的原因是能够与 GAL4 的二聚化域（第 58～97 位氨基酸之间）的一个区域结合。由于 GAL11（或 GAL11P）是全酶的一部分，GAL11P 与 GAL4 间的新结合可将全酶召集到 GAL4 应答的启动子上，如图 12.16 所示。之所以称这种结合为“新的”，是因为 GAL11P 参与结合的部位在正常情况下是无功能的，GAL4 参与的部位在二聚化域而不是激活域。正常情况下这两个蛋白质区域不会发生任何形式的结合。

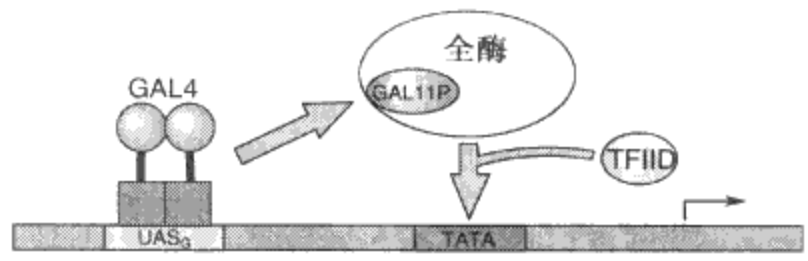


图 12.16　GAL11P 通过 GAL4 二聚化域召集聚合酶全酶的模型。在聚合酶全酶中，GAL4 的二聚化域（黄色箭头）结合 GAL11P（紫色），使聚合酶全酶与 TFIID 一起结合在启动子上，激活基因转录。

为验证 GAL4 第 58～97 位氨基酸之间的区域是 GAL11P 转录激活的关键，Ptashne 及同事进行了以下实验。他们克隆构建了编码 GAL4 的 58～97 位氨基酸区段和 LexA 的 DNA 结合域的融合蛋白质粒，

并将其与另外两种质粒一起导入酵母细胞。一个编码 GAL11 或编码 GAL11P，另一个带有 LexA 的两个结合位点，位于 GAL1 启动子上游，驱动下游大肠杆菌 *lacZ* 报告基因。图 12.17 总结了该实验结果。当全酶含有野生型 GAL11 时，LexA-GAL4（58～97）蛋白无激活因子功能（图 12.17a），当全酶含有 GAL11P 时，LexA-GAL4（58～97）蛋白有激活因子作用（图 12.17b）。

如果激活作用确实是由于 LexA-GAL4（58～97）和 GAL11P 之间的相互作用引起，那么将 LexA 的 DNA 结合域与 GAL11 融合也应该激活转录，如图 12.17c 所示。实际上，这样构建的载体确实有激活效应，与假设一致。因为 LexA 已经与 GAL11 共价连接了，无须再发生新的像 LexA-GAL4 和 GAL11P 之间的互作。

对这些结果最简单的解释就是，至少在这一体系中，转录激活可通过召集聚合酶全酶，而非单个通用转录因子来实现。有一种可能但不常见的情况，即 GAL11 是一种特殊蛋白，它的召集作用引起前起始复合物的逐步组装。我们更倾向于认为激活因子和全酶中任意组分间的相互作用都能召集全酶，从而激活转录。Ptashne 及同事承认 TFIID 是前起始复合物的必要组分，但显然不是酵母全酶的组成部分。他们提议，在该实验中 TFIID 可能与全酶协同结合到启动子上。

至少有两个证据表明全酶并不作为整体被召集。首先，David Stillman 及同事做了各种因子结合酵母 *HO* 启动子区的动力学实验。结果表明，全酶的一部分，即中介物，早在 RNA 聚合酶 II 全酶结合的 G_1 期就与启动子结合了。因此，至少对酵母的这一启动子来说，全酶肯定不是以整体结合的。

其次，Roger Kornberg 及同事分析，如果全酶以一个整体与启动子结合，那么在细胞中全酶各组分的含量应大致相等。他们知道测定细胞中蛋白质含量很困难，不能通过检测 mRNA 水平来获得结果，因为在转录后的加工事件有许多变化，如 mRNA 降解及核输出等。的确，mRNA 的浓度与其相应蛋白质产物的浓度可能与预期值偏差 20～30 倍。可以通过双向凝胶电泳分离蛋白质，并通过质谱仪确定其浓度（第 24 章），但该方法对于在体内浓度很低的转录因子类型的蛋白质不够灵敏。

因此，Kornberg 及同事选择了一种兼具高灵敏度和准确性的方法。他们先用基因克隆技术，将“TAP”标签附加在编码聚合酶 II 全酶不同组分的基因上，包括 RNA 聚合酶 II、中介物及 5 种通用转录因子。TAP 标签包含葡萄球菌蛋白 A（第 4 章）的可结合抗体 IgG 的区域。然后对携带编码 TAP 标

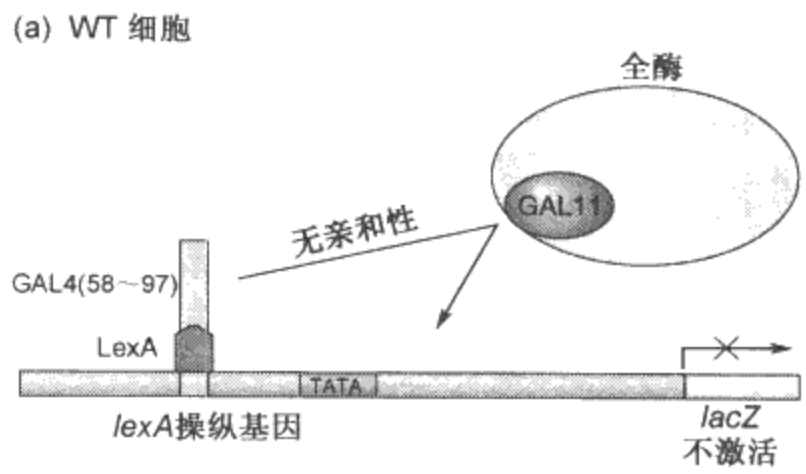


图 12.17　GAL11P 和 GAL11-LexA 的激活作用。Ptashne 等用一个包含 *lexA* 操纵基因（驱动 *lacZ* 转录，位于启动子上游 50bp 处）的质粒和下列质粒转染真核细胞：(a) 一个编码 GAL4 58～97位氨基酸和 LexA 的 DNA 结合域的质粒，以及编码野生型 GAL11 的质粒；(b) 一个编码 GAL4 的 58～97 位氨基酸和 LexA 的 DNA 结合域的质粒，以及编码 GAL11P 的质粒；(c) 一个编码 GAL11 和 LexA 的 DNA 结合域的质粒。他们检测了 *lacZ* 的产物——β-半乳糖苷酶的产量。结果：(a) GAL4（58～97）区不与 GAL11 相互作用，因此不发生转录；(b) GAL4（58～97）区与 GAL11P 结合，召集聚合酶全酶到启动子上，所以激活转录；(c) LexA-GAL11 融合蛋白能结合到 *lexA* 操纵基因上，召集聚合酶全酶到启动子上，从而激活转录。（*Source*：Adapted from Barberis A., J. Pearlberg, N. Simkovich, S. Farrell, P. Resnagle, C. Bamdad, G. Sigal, and M. Ptashne, with a component of the polymerase II holoenzyme suffices for gene activation. *Cell* 81：365，1995.）

记蛋白基因的酵母菌株提取物进行斑点印迹，用抗过氧化物酶抗体进行探测。印迹膜中蛋白质上的TAP标签与抗体结合，它们再与后加入的过氧化物酶结合，使氧化物酶底物变成一种可被成像仪检测的化学发光物质（第5章）。

胶片上斑点的强度与印迹膜中TAP标记蛋白的浓度正相关。对各种提取物进行一系列稀释，根据标准蛋白GST-TAP的一个已知含量的印迹结果，可将斑点的不同强度换算成细胞中各蛋白质的浓度。图12.18显示了实验结果。很明显，无TAP标记的野生型样品的泳道背景近似零，这对定量的准确性非常重要。从图中还看到，RNA聚合酶II的含量远比Med8（中介物的亚基之一）多。定量结果（图12.18b）显示，Rpb3含量是中介物或TFIIH亚基的5～6倍。表12.1列举了TFIIF、TFIIE、TFIIB、TFIID以及图12.18中所列蛋白质的定量值。表中又一次显示RNA聚合酶比其他因子的量更高，而其中4个通用转录因子的量比中介物或TFIIH更大。

因为实验中未发现大致等量的全酶组分，所以全酶以整体结合大多数启动子不太可能。尽管如此，某些启动子上仍有可能以整体形式召集全酶。

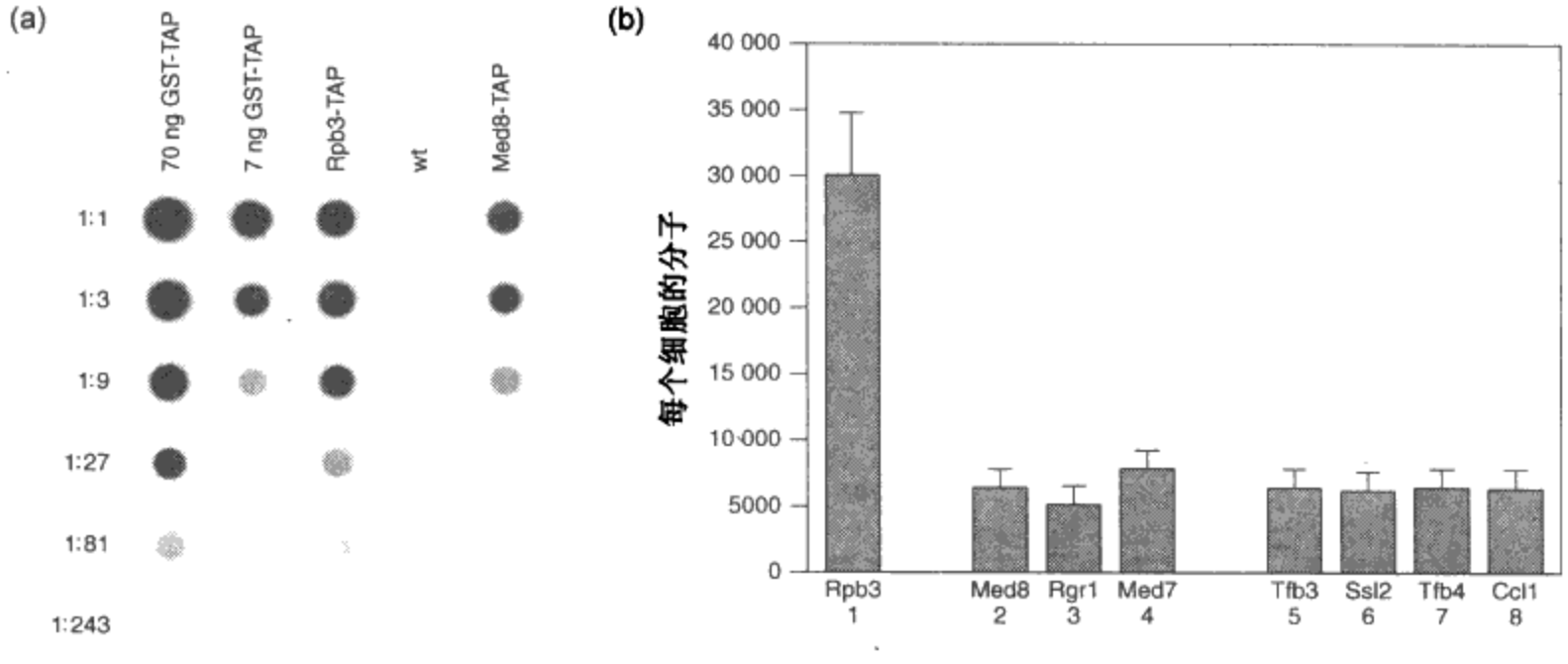


图12.18　斑点杂交实验测定聚合酶全酶亚基的浓度。（a）斑点印迹结果。Kornberg及同事对携带编码全酶亚基（被TAP序列标记的）嵌合基因的酵母系列稀释提取物进行了斑点印迹。TAP序列含有两个葡萄球菌蛋白A序列，可与IgG免疫球蛋白结合。用IgG免疫球蛋白（兔抗过氧化物酶IgG）与斑点印迹上的TAP序列反应。随后用过氧化物酶与底物反应，产生可用成像仪检测的化学发光物质来检测IgG。左边列出了稀释比例。1～2列显示两种GST-TAP浓度的系列稀释结果。3～5列分别包含TAP标记的Rpb3、无TAP标签的野生型细胞以及TAP标记的Med8提取物的系列稀释结果。（b）由斑点印迹测得的各组分在细胞中的浓度。1为Rpb3；2～4为中介物的3个亚基；5～8为TFIIH的4个亚基。（*Source*：*Journal of Biological Chemistry* by Borggrefe et al. Copyright 2001 by Am. Soc. For Biochemistry & Molecular Biol. Reproduced with permission of Am. Soc For Biochemistry & Molecular Biol. in the format Textbook via Copyright Clearance Center.）

表12.1　每个酵母细胞中所选蛋白质分子的拷贝数

蛋白质	拷贝数/个细胞	蛋白质	拷贝数/个细胞
RNA聚合酶II（Rpb3）	30 000	TFIID（TBP）	20 000
TFIIF（Tfg2）	24 000	中介物（Med8）	6 000
TFIIE（Tfa2）	24 000	TFIIH（Tfb3）	6 000
TFIIB（Sua7）	20 000		

Source：Borggrefe，T.，R. Davis，A. Bareket-Samish，and R. D. Kornberg，Quantitation of the RNA polymerase II transcription machinery in yeast. *Journal of Biological Chemistry* 276（2001）：47150-53，tll. Reprinted with permission.

小结 至少在酵母的某些启动子上，转录激活通过召集全酶而不是一次一个组分地组装进行。然而，其他证据也表明召集全酶的方式并不常见。

12.5 激活因子间的相互作用

我们已经了解到不同类型转录因子之间的主要作用方式。显然，通用转录因子之间必须相互作用以形成起始复合物前体，但转录激活因子和通用转录因子间也会相互作用。例如，我们讨论过 GAL4 等转录激活因子与 TFIID、TFIIB 及其他通用转录因子间的相互作用。此外，激活因子间也会相互作用以激活同一个基因。这种作用有两种发生方式：一是单体相互作用形成蛋白质二聚体，促进对单一 DNA 靶位点的结合；另一种是结合在不同 DNA 位点的特异激活因子可以协同激活一个基因。

二聚化作用

我们已经提到在 DNA 结合蛋白中，蛋白质单体间有许多不同的作用方式。在第 9 章我们讨论了螺旋-转角-螺旋蛋白 λ 阻遏物，该蛋白质的两个单体间的相互作用使它们的识别螺旋刚好处在与两个 DNA 大沟相互作用的位置，而另一个螺旋转开。识别螺旋呈反向平行排列，使它们能识别具有回文序列的 DNA 靶位点的两部分。在本章前面我们已经讨论了 GAL4 蛋白螺旋圈的二聚化域、bZIP 蛋白与之类似的亮氨酸拉链及肾上腺糖皮质激素受体的环化二聚化域等内容。

在第 9 章，我们介绍了蛋白质二聚体相对于单体对 DNA 结合的优势，蛋白质与 DNA 的结合力跟结合自由能的平方正相关。因为自由能取决于蛋白质和 DNA 接触的数量，蛋白质二聚体取代蛋白质单体使接触次数加倍，而蛋白质和 DNA 之间的亲和力增加至原来的 4 倍。这对于大多数激活因子都必须在低浓度下行使功能具有重要意义。事实上大多数 DNA 结合蛋白都以二聚体存在就证明了它的优势。让我们看一个众所周知的异源二聚体（heterodimer）（由两个不同单体组成）的实例，Jun 和 Fos。

Jun-Fos 二聚体 Jun 和 Fos 是 *jun* 和 *fos* 基因（原癌基因，在某些肿瘤发生中起作用）的产物。Jun 和 Fos 蛋白的 DNA 结合域属于 bZIP 家族。Jun 二聚体是激活因子 AP1 的一种形式，但是 Jun 和 Fos 异源二聚体结合 DNA 比 Jun 二聚体更加牢固（Fos 二聚体根本不结合 DNA），所以大部分 AP1 的自然形态可能是 Jun-Fos 异源二聚体。

T. Kouzarides 和 E. Ziff 证明，Jun 和 Fos 以异源二聚体结合 DNA 比各自以单体形式结合更有效，同时每个单体的 bZIP 域对于结合都非常重要。他们对两种单体的蛋白及其 bZIP 域进行了表达（在体外转录相应基因然后翻译其 mRNA）。将这些蛋白的不同组合与含有 AP-1 结合位点（TGACTCA）的寡核苷酸混合。AP-1 结合位点是 **TPA** 应答元件 **TRE**（TPA response element），对包括 TPA 在内的一类复合物产生应答。TPA（12-*O*-十四烷酰佛波醇-13-乙酸酯，12-*O*-tetradecanoylphorbol-13-acetate）这类佛波酯化合物通过 TRE 激活基因转录而促使细胞分裂。最后，利用凝胶阻滞实验检测蛋白质与 DNA 的结合情况。

图 12.19 是实验结果。在低浓度下，Fos 和 Jun 单独都不能结合 DNA，但两个蛋白质的复合物可以结合 DNA。同样，包含 bZIP 域的 Fos 和 Jun 核心部分必须联合起来共同结合 DNA（他人研究发现，Jun 可以在高浓度时独自结合 DNA）。

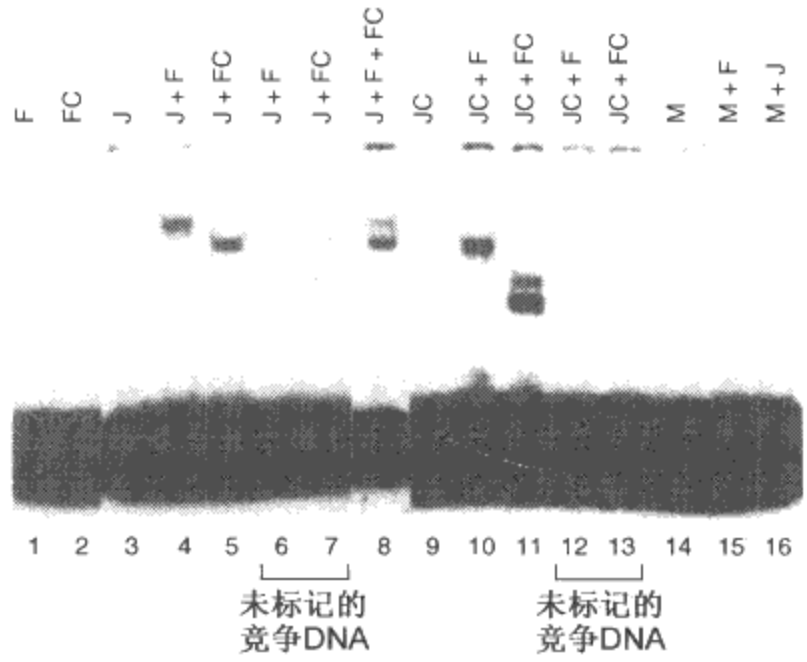


图 12.19 Fos 和 Jun 在 AP-1 结合位点与 DNA 的结合。 Kouzarides 和 Ziff 将 Fos、Jun 及 Jun 和 Fos 的 bZIP 结构域（又称 Fos 核心和 Jun 核心）与标记的带有 AP-1 结合位点的 DNA 混合，用凝胶阻滞实验检测蛋白质和 DNA 的结合。各泳道顶部的字母表示加入的蛋白质（F＝Fos，J＝Jun，FC＝Fos 核心，JC＝Jun 核心，M＝Myc，另一个 bZIP 蛋白）。泳道 6、7、12、13 含有过量无标记 DNA 竞争物，阻止蛋白质与标记 DNA 结合。注意 Fos/Fos 核心与 Jun/Jun 核心之间的二聚体能结合 DNA，但各单体不能结合 DNA。它们之间的相互作用是特异的，因为 Myc 与其不能发生相互作用。因此，Fos-Jun 异源二聚体比其同源二聚体更能有效地结合 AP-1 位点。(*Source*: Kouzarides T. and E. Ziff, The role of the leucine zipper in the Fos-Jun interaction. *Nature* 366 (15 Dec 1988) p. 649, f. 3a. Copyright © Macmillan Magazines Ltd.)

该实验也证明 Fos-Jun 复合物是二聚体，而不是四聚体。第 8 泳道含有 Jun、Fos 及 Fos 核心部分。从泳道 4 和 5 可以看出，Jun-Fos 和 Jun-Fos 核心与 DNA 结合产生不同迁移率的两个条带。因此，我们预计 Jun、Fos 和 Fos 核心的混合物除产生以上两种条带（形成二聚体）外，还可能显示第三条带（Jun_2-Fos-Fos 核心的四聚体）。但没有第三条带出现，意味着这些蛋白质只形成二聚体。这一结果并非说明转录激活因子只以二聚体形式结合 DNA，肿瘤抑制物 p53 就是通过四聚体行使转录激活作用的。

同样的二聚化和 DNA 结合反应会发生在 Fos/Jun 与其他 bZIP 蛋白（如 Myc）之间吗？显然不会，因为无论是 Fos 与 Myc 或 Myc 与 Jun 的混合物，都不能结合具有 TRE 的 DNA 序列（泳道 15 和 16）。因此，尽管 bZIP 蛋白间具有相似的结构，但它们与其对应的 DNA 靶位点间的相互作用却是特异的。

小结 二聚化的巨大优势在于其增加了激活因子和 DNA 的亲和力。有些激活因子形成同源二聚体，有些则形成更高效的异源二聚体，如 Jun 和 Fos。

远程作用

我们知道原核生物和真核生物的增强子即使远离启动子也能激活相应基因的转录。这种远程调控作用是如何发生的呢？在第 9 章我们知道，实验证据表明 DNA 的两个远距离位点之间环凸使与之结合的两个蛋白质相互作用。这一机制也适用于真核生物的增强子。

目前有如下几种较有说服力的增强子远距离作用假说（图 12.20）：①激活因子与增强子结合，通过超螺旋使整个 DNA 双链的拓扑结构或形状发生

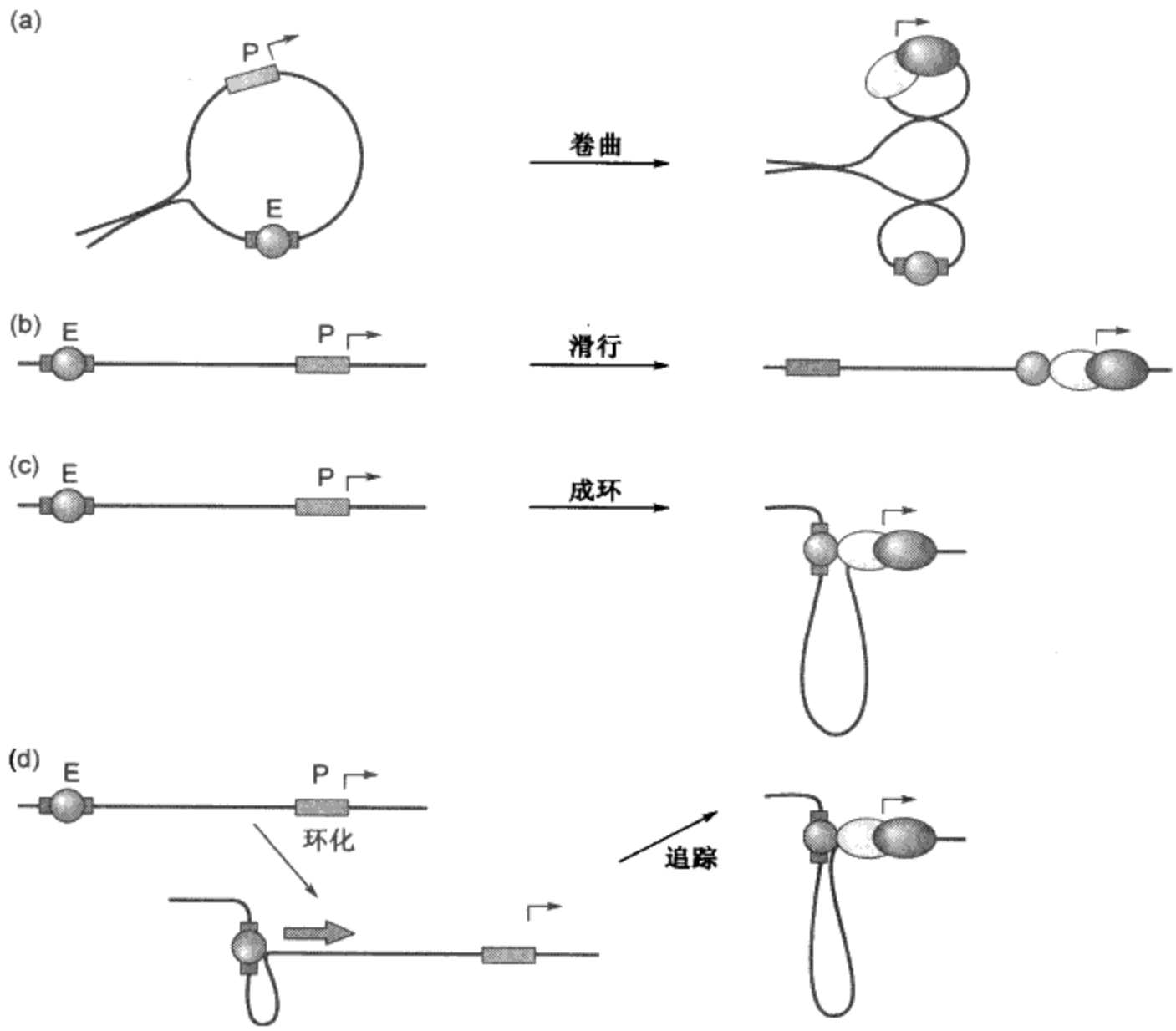


图 12.20 增强子作用的 4 种假说。（a）拓扑结构的改变。增强子（E，蓝色）和启动子（P，橙色）定位在同一 DNA 环上。基因特异性转录因子（绿色）与增强子结合形成超螺旋结构，使通用转录因子（黄色）和 RNA 聚合酶（红色）结合到启动子上。（b）滑动。转录因子与增强子结合并沿 DNA 滑动到启动子附近，协助通用转录因子和聚合酶结合到启动子上。（c）环化。转录因子结合到增强子上，通过使增强子与启动子间的 DNA 环凸而结合启动子，并促使通用转录因子和聚合酶结合到启动子上。（d）促追踪（facilitated tracking）。转录因子与增强子结合，使下游一段短 DNA 环化。随着环的增大使转录因子沿 DNA 追踪直至找到启动子，然后帮助通用转录因子和 RNA 聚合酶结合启动子。

改变，由此使启动子向通用转录因子开放；②激活因子与增强子结合，然后沿 DNA 滑动直至遇到启动子，并通过与启动子的直接作用来激活转录；③激活因子与增强子结合，通过增强子与启动子间的 DNA 环凸使激活因子与启动子上的蛋白质产生作用，激活转录；④激活因子结合增强子后，在其下游 DNA 链上产生环凸，通过增大此环，激活因子向着启动子方向移动，到达启动子并与之相互作用激活转录。

注意，前两种模型要求增强子和启动子在同一 DNA 分子上。一个 DNA 分子构型的变化不会影响另一个分子的转录，与一个 DNA 分子结合的激活因子也不能滑动到第二个含启动子的 DNA 分子上。而第三种模型只要求增强子与启动子彼此相互靠近，不一定在一个分子上。这是因为环化模型的本质不是 DNA 分子本身环化，而是结合在两个远距离位点上的蛋白质间的相互作用。原则上讲，如果蛋白质是结合于不同 DNA 分子的两个位点，只要 DNA 分子以某种方式连在一起不漂移开，并阻止结合的蛋白质之间的相互作用，那么第三种模型就很有效。图 12.21a 显示这一模型的发生过程。

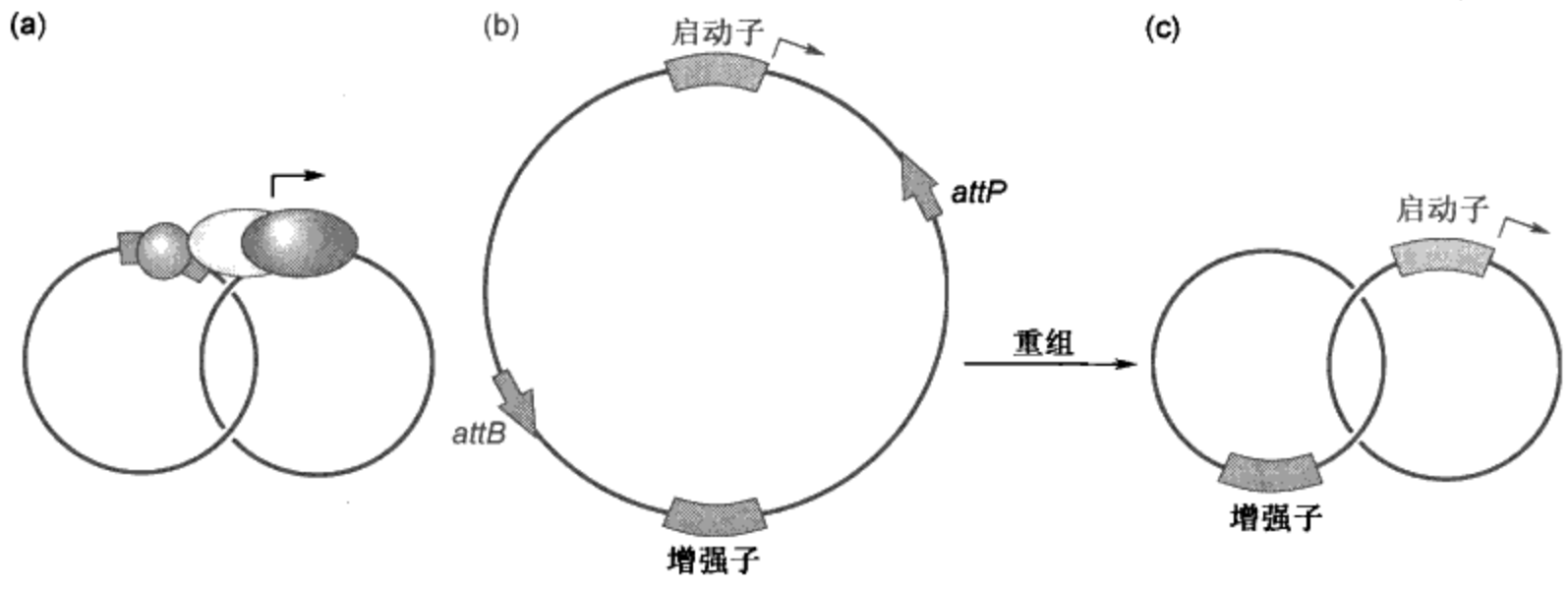


图 12.21　增强子和启动子分别位于连环体的不同质粒上。（a）基因特异转录因子（绿色）与通用转录因子（黄色）和聚合酶（红色）之间可能的相互作用模式。特异性转录因子结合在连环体一个质粒的增强子（蓝色）上，通用转录因子和聚合酶结合在连环体的另一个质粒的启动子（图中看不见）上。（b）连环体的产生。增强子和启动子定位于质粒的两个相对位置上，由 λ 噬菌体 *attB* 和 *attP* 位点将两者隔开。（c）在体外两个 *att* 位点发生重组，产生两个质粒分别含有增强子和启动子的连环体。

因此，如果能使一个 DNA 分子的增强子和另一个分子的启动子以**连环体**（catenane，圆环连成链状）形式连在一起，可验证以上假设。如果其中的增强子仍发挥作用，就能排除前两种假说。Marietta Dunaway 和 Peter Dröge 做了这个实验。他们用非洲爪蟾 rRNA 启动子，其两边分别连接 rRNA 微小基因和 rRNA 增强子，并在中间加入 λ 噬菌体的整合位点 *attP* 和 *attB* 构建了一个质粒，如图 12.21b 所示。*attP* 和 *attB* 都是位点特异性重组的靶位点，因此将它们放在同一个分子上，可重组产生连环体，如图 12.21c 所示。

最后，研究者将不同的质粒组合注入非洲爪蟾卵母细胞，通过定量 S1 作图法测定转录情况。注入的质粒分别是：连环体质粒、含有增强子和启动子的未重组质粒、只含增强子或启动子的质粒。在 S1 定量分析中需要一个参照质粒来校正卵母细胞间的差异。这里的参照质粒包含一个 rRNA 微小基因（ϕ52），有 52bp 的插入片段，而测试质粒（ϕ40）的 rRNA 微小基因都含有 40bp 的插入片段。用这两种微小基因分别制备了标记探针，如果两个基因都发生转录，应该看到相差 12nt 的两个信号。我们最想知道这两种信号的强度比值，它表示测试质粒相对于参照质粒的转录效率（参照质粒的效率在各实验中应保持一致）。

图 12.22a 中，“a”的泳道显示测试质粒的结果，“b”泳道显示参照质粒的结果。图 12.22b 显示各泳道转录产物对应的质粒。注意，图 12.22a 中每组实验的 a 泳道和 b 泳道所用的质粒相同，只有探针不同。结果如下：泳道 1 表示质粒只含启动子时，测试质粒的信号比参照质粒弱。这是因为测试探针的特异放射性比参照探针低。泳道 2 表明紧挨着启动子的增强子（正常位置）极大地提高了测试质粒的转录水平，其信号比参照质粒强得多。泳道 3 显示当增强子与启动子在质粒上相对放置时，虽有转录活性，但效率不高。泳道 4 显示的信息最重要，表明当增强子和启动子分别位于连环体的两个不同质粒上时，增强子仍然起作用。泳道 5 证明当增强子和启动子位于两个不同质粒但未形成连环体时，增强子无效。泳道 6 显示在泳道 4 观察到的增强效应不是由少量非重组质粒污染造成。因为在泳道 6

中加入5%的非重组质粒时，测试质粒信号无显著　　增加。

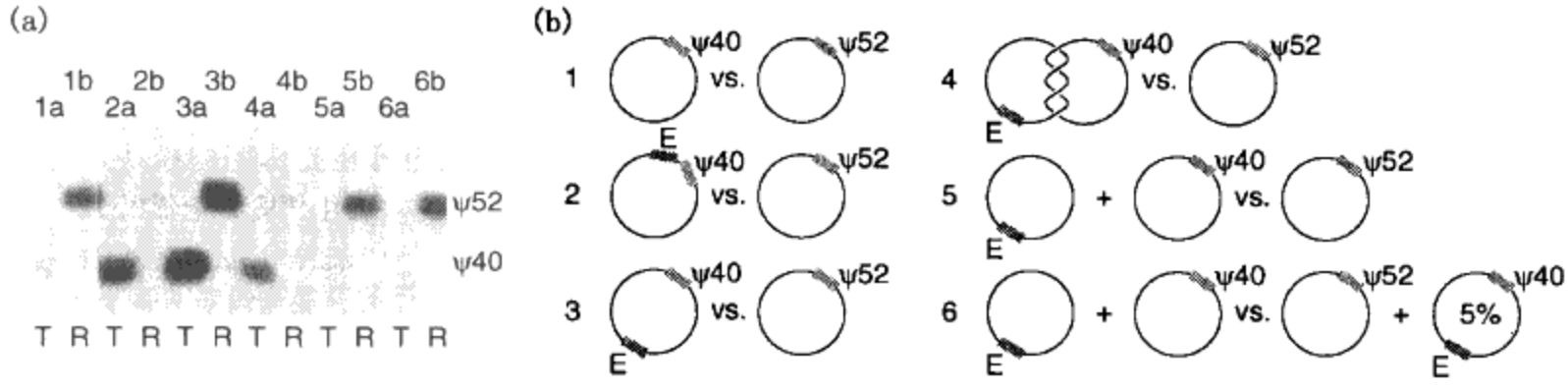


图 12.22　连环体实验结果。Dunaway 和 Dröge 将混合质粒注入非洲爪蟾卵母细胞并通过定量 S1 作图法测定转录效率。每组实验注入参照质粒和一种测试质粒的混合物，分别用不同的探针检测转录情况。(a) 实验结果。测试质粒（T）和参照质粒（R）的结果见泳道 a 和 b。每个实验注入的不同质粒在（b）图给出。如，1a 和 1b 泳道所用质粒标记为 1。左边标记 ϕ40（或 ϕ40＋另一质粒）为测试质粒。右边标记 ϕ52 的是参照质粒。名称中的 40 和 52 表示插入片段大小，以示区别。两种质粒同时注入，然后分别用测试探针（泳道 1a）或参照探针（泳道 1b）检测。泳道 4a 和 4b 显示分别含启动子和增强子于两个质粒的连环体的转录效率比只含启动子的质粒（泳道 1a 和 1b）高。显然泳道 4a 和 4b 的转录信号比值比泳道 1a 和 1b 高得多。(*Source*：Adapted from Dunaway M. and P. Dröge, Transactivation of the *Xenopus* rRNA gene promoter by its enhancer. *Nature* 341 (19 Oct 1989) p. 658, f. 2a. Copyright © Macmillan Magazines Ltd.)

综上所述，关于增强子的功能得到以下两点结论：增强子不必和启动子位于同一个 DNA 分子上，但它必须在空间上与启动子足够靠近，以便结合在增强子和启动子上的蛋白质可以相互作用。这很难与超螺旋模型及滑动模型相容（图 12.20a 和 b），但与 DNA 环化和促追踪模型相一致（图 12.20c 和 d）。在连环体结构中不需要环化或追踪，因为增强子和启动子位于不同的分子上，而其蛋白质间的相互作用无须 DNA 环化作用，如图 12.21a 所示。

小结　增强子作用的本质是结合在增强子上的激活因子与结合在启动子上的通用转录因子及 RNA 聚合酶之间发生的相互作用。这可能以增强子和启动子之间 DNA 的环凸介导，至少在理论上是这样。DNA 环凸可以将结合在不同增强子上的激活因子拉至足够靠近启动子的位置，然后协同激活基因转录。

复合增强子

许多基因上有不止一个增强子，以此应答多重刺激。例如，金属硫蛋白基因编码的蛋白质显然是帮助真核生物应对重金属毒害的，可被不同的效应物激活，如图 12.23 所示。因此，每个结合在多重增强子上的激活因子一定对起始复合物前体在启动子上的组装起作用，可能是通过将增强子和启动子间的 DNA 成环外突而实现。

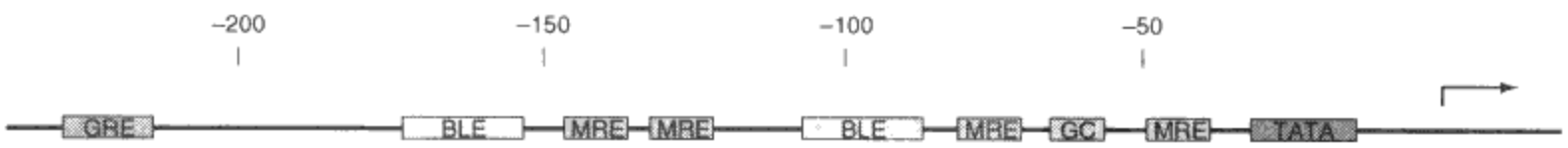


图 12.23　人金属硫蛋白基因的调控区。位于转录起始位点＋1 上游，沿 3′→5′的方向分别为：TATA 盒、金属响应元件（MRE)-在应答重金属时激活基因转录、GC 框-转录激活因子 Sp1 的响应元件、另一个 MRE 序列、本底水平增强子（BLE)-转录激活因子 AP-1 的响应元件、两个串联的 MRE 序列、另一个 BLE 序列、糖皮质激素响应元件（GRE)-可被糖皮质激素和受体组成的激活因子刺激转录。

对多个激活因子结合位点同时调节一个基因的发现，正在改变我们对“增强子”一词的定义。增强子最初被定义为一种非启动子 DNA 元件，与至少一种增强子结合蛋白一起刺激邻近基因的转录。因此，图 12.23 中所示的金属硫蛋白基因位于 TATA 盒上游的调控区应包含多个增强子。但现在增强子的定义已经朝着涵盖启动子以外相邻的整个调控区方向演变。因此，金属硫蛋白基因的全部调控区均可视为一个增强子，如 BLE 仅是增强子中的一个元件。尽管采用了这种新的定义，我们仍可认为某些基因由多增强子控制。例如，果蝇黄（*yellow*）和白（*white*）基因（在本章后面将讨论）就是由 3

个增强子控制——3 簇相邻的激活因子结合位点。

与多个激活因子相互作用的增强子允许对基因表达的精细调控。不同激活因子的组合对不同细胞中的给定基因产生不同水平的表达。事实上，在基因附近各种增强子元件的出现或缺失使人联想到二进制代码，元件出现为“开”，缺失则为“关”。当然，必须由激活因子来操纵开关。由于多增强子元件之间是协同作用的，因此并不表现为简单的加性效应。

另一个可以用来描述多重激活因子作用的巧妙比喻是**组合代码**（combinatorial code）。特定细胞特定时间下，所有激活因子的集合构成组合代码。如果一个基因带有一组增强子元件，每个增强子可应答一至多个激活因子，就可以阅读此代码。其结果是基因以适当的水平进行表达。

Eric Davidson 及同事以海胆 *Endo16* 基因提供了一个绝好的多增强子例子。*Endo16* 基因在动物早期胚胎的营养板（指后期发育中产生内皮组织，包括内脏的细胞群）发育中很活跃。他们最初测试了 *Endo16* 基因 5′-侧翼区结合核蛋白的能力。结果发现，这样的结合区很多，排布成 6 个模块，如图 12.24 所示。

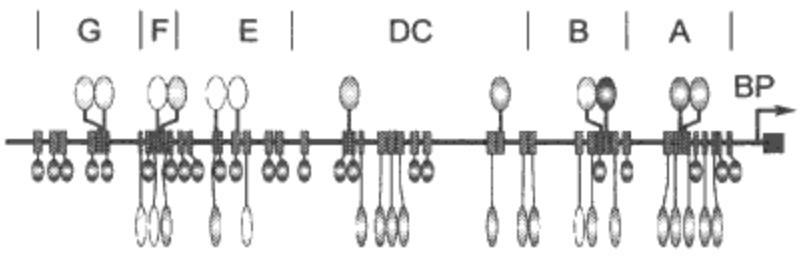


图 12.24　海胆 *Endo16* 基因多增强子模块的排布方式。大的彩色椭圆表示激活因子（译者注：原文为椭圆和长方形），小的蓝色椭圆表示结构转录因子，两者都结合到增强子元件上（红色框）。增强子呈簇状或模块排布，分别标记为 G、F、E、DC、B 和 A。长垂直线代表不同模块的区分位置。BP 代表“基本启动子”。（*Source*: Adapted from Romano, L. A. and G. A. Wray, Conversation of *Endo 16* expression in sea urchins despite evolutionary divergence in both cis and trans-acting components of transcriptional regulation, *Development* 130 (17): 4189, 2003.）

怎样知道与核蛋白结合的这些模块中哪些真正与基因激活有关呢？Chiou-Hwa-Yuh 和 Davidson 将这些模块单独或组合起来并与 CAT 报告基因连接，然后导入海胆卵细胞，观察在胚胎发育中报告基因的表达模式。他们发现，报告基因在胚胎组织表达的部位和时间取决于导入的不同模块的精确组合形式。这些模块（内含增强子）对不均一分布在发育胚胎中的激活因子产生应答。

也许所有元件在体外都能独立作用，但在体内却是组织化的，如图 12.25 所示。模块 A 可能是唯一直接与基本转录组件相互作用的模块，其他模块都需通过模块 A 发挥作用。某些上游模块（如 B 和 G）通过 A 协同作用以激活 *Endo16* 基因在内皮细胞发育中的转录。而另一些模块（如 DC、E 和 F）也通过 A 协同阻抑非内皮细胞中 *Endo16* 基因的转录（模块 E、F 在外胚层细胞中起激活作用，而模块 DC 则在成骨间质细胞中起作用）。在该实验中最吸引人的是加入 LiCl 产生的效果。LiCl 是一种致畸剂（teratogen）（畸形发育的诱导剂）。在海胆胚胎中，LiCl 促使内胚层过度发育从而抑制附近外胚层的发育。Davidson 及同事证明，LiCl 使模块 DC、E、F 通过模块 A 的作用发生了逆转，使原来抑制 *Endo16* 活性变成激活转录的作用。因此，LiCl 使原外胚层细胞中本该关闭的基因被开启，从而过量产生内胚层细胞，使原外胚层细胞发育成内胚层。

小结　复合增强子能使一个特定基因对各种激活因子的不同组合产生应答。这种排布格局使细胞能够对基因在不同组织或不同发育阶段产生精细调控。

结构转录因子

我们已经知道，将激活因子和通用转录因子拉到一起的 DNA 环凸机制非常实用，尤其是对于结合到相距几百个碱基的不同 DNA 元件上的蛋白质，因为 DNA 的柔韧性足以产生这种弯曲。但是，许多增强子距其控制的启动子较近，而短小的 DNA 更像僵硬的短棒，不像“柔韧的带子”，在此情况下，DNA 环凸不会自然发生。

那么，紧密结合在一小段 DNA 上的激活因子和转录因子如何相互作用来刺激转录呢？如果中间有其他因子干涉促使 DNA 分子进一步弯曲，那它们仍能相互靠近。我们已经遇到几个关于**结构转录因子**（architectural transcription factor）的实例，其主要作用是改变 DNA 调控区的形状，以便其他蛋白质间能相互作用刺激转录。Rudolf Grosschedl 及同事首先提供了一个真核生物结构转录因子的例子，人 T 细胞受体 α 链（TCRα）基因的调控区，含有 3 个位于转录起点上游 112bp 以内的增强子，分别为激活因子 Ets-1、LEF-1 和 CREB 的结合位点（图 12.26）。

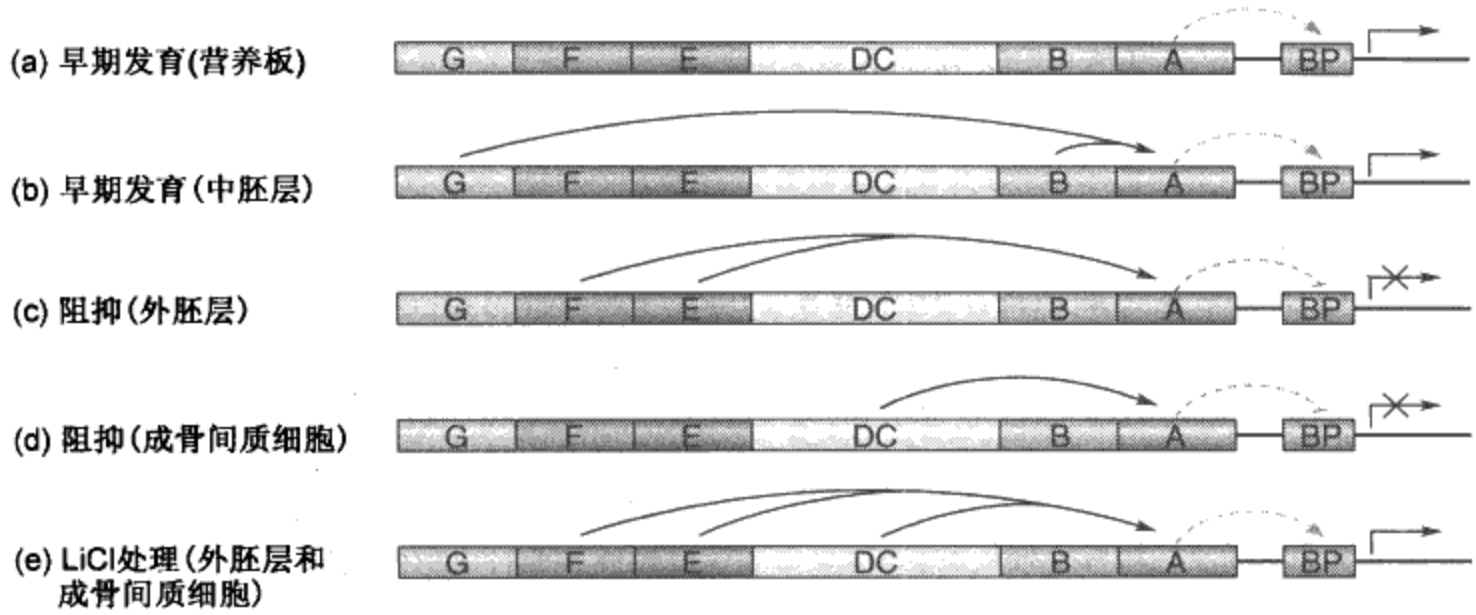


图 12.25　海胆 *Endo16* 基因中模块 A 的综合功能。（a）在早期发育中，A 是唯一开启的模块，它通过基本启动子（BP）刺激转录，如虚线箭头（绿色）所示；（b）在中胚层的晚期发育中，模块 A 通过整合模块 G 和 B 的作用，共同刺激转录；（c）在外胚层晚期发育中，模块 A 通过整合模块 F 和 E 的作用，阻抑基因转录，如红色阻抑虚线所示；（d）在成骨间质细胞晚期发育中，模块 A 整合模块 DC 的作用，抑制基因转录；（e）LiCl 处理能够逆转来模块 A 对外胚层和成骨间质细胞的阻抑信号，刺激基因转录。（*Source*：Adapted from Yuch，C.-H.，B. Hamid，and E. H. Davidson，Genomic *cis* regulatory logic：Experimental and computational analysis of a sea urchin gene. *Science*，279：1897，1998.）

图 12.26　人类 T 细胞受体 α 链（TCRα）基因的调控区。在转录起点上游 112bp 以内有 3 个增强子元件，分别与 Ets-1、LEF-1 及 CREB 结合。这 3 个增强子以其结合的转录因子而不是自己的名字来区分。

LEF-1　是淋巴增强子结合因子，它结合图 12.26 中间的增强子，帮助激活 TCRα 基因。然而，Grosschedl 等前期的研究结果表明，LEF-1 自身不能激活 TCRα 基因。那么它如何起作用呢？Grosschedl 及合作者证实，LEF-1 通过与增强子的小沟结合将 DNA 分子弯曲 130°。

他们用两种方法证实了 LEF-1 与增强子小沟的结合。首先，对增强子的 6 个腺嘌呤的 N3（位于 DNA 小沟）甲基化可干扰增强子的功能。然后，将这 6 对 A-T 换成 6 对 I-C（它们的小沟看起来相同），不减弱增强子活性。这与 Stark 和 Hawley 证明 TBP 和 TATA 盒小沟结合的策略（第 11 章）相同。

Grosschedl 及同事采用与 Wu 和 Crothers 相同的电泳检测法（为显示 CAP 蛋白可使乳糖操纵子弯曲，第 7 章），证明了 LEF-1 可以使 DNA 弯曲。将 LEF-1 结合位点置于线性 DNA 片段的不同位置，并与 LEF-1 结合，然后检测其电泳迁移率。当结合位点在 DNA 片段中部时，泳动受到很大阻滞，这表明此时 LEF-1 使 DNA 产生了显著的弯曲。

他们还进一步证实了 DNA 分子的弯曲是位于 LEF-1 上 **HMG 域**（HMG domain）所致。**HMG 蛋白**是一类具有高泳动率（由此得名 high mobility group，HMG）的小核蛋白。为证明 LEF-1 的 HMG 域的重要性，研究人员制备了只含 HMG 域的纯多肽，并显示此肽与全长蛋白质一样能使 DNA 分子弯曲 130°。将迁移速率曲线外推至最大速率点［此时致弯元件（bend-inducingelement）正好在 DNA 片段的末端］后发现弯曲发生在 LEF-1 的结合位点上。由于 LEF-1 本身不能增强转录，因此可能通过使 DNA 弯曲而间接发挥作用。这样就可能使其他激活因子结合启动子的基本转录装置从而增强转录。

Tom Maniatis 及同事在人 β-干扰素（IFN-β）基因的调控区（含有一个靠近启动子的增强子）观察到类似的现象。该增强子可结合多个激活因子，包括 HMG 蛋白家族中不同于 LEF-1 的 HMG I（Y）。HMG I（Y）蛋白家族成员不使 DNA 弯曲，但调控富 A-T 的 DNA 的自然弯曲，使其能与其他蛋白质相互作用。Maniatis 及同事用凝胶电泳实验显示，HMG I（Y）跟**核因子 NF-κB**（nuclear factor kappa B）激活因子在与富含 A-T 的靶位点协同结合时，未使 DNA 弯曲。

HMG I（Y）对病毒诱导表达的 IFN-β 基因是必需的，但无其他激活因子时并不发挥激活因子功能。Maniatis 及同事证明，HMG I（Y）对其他两个激活因子结合增强子是必需的，一个是 ATF-2（激活转录因子 2，activation-transcription factor 2），另一个是 NF-κB。HMG I（Y）的作用可能是重塑增强子 DNA，以便帮助其他激活因子在**增强体**（enhanceosome）内互相作用，如图 12.27 所示。这可以解释一个明显的悖论：IFN-β 是一个抗病毒蛋白，其编码基因在病毒侵染细胞中被诱导，然而 IFN-β 基因还有一个与 NF-κB 结合的增强子，**NF-κB** 除了被病毒侵染诱导外，还对多种信号及胁迫产生应答。那么为什么其他信号和胁迫条件不能激活 IFN-β 基因的表达呢？答案就是激活此基因需要整个增强体组装成合适的形状，而这就要求所有激活因子在病毒侵染的初期就同时存在。

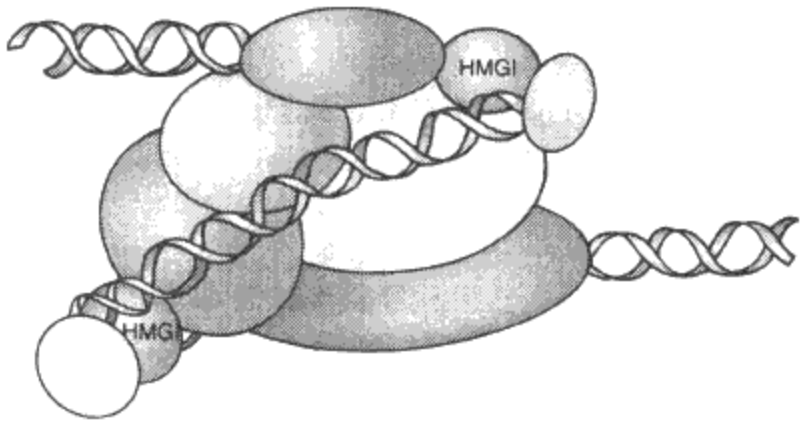


图 12.27　示意性的增强体模型。激活因子和通用转录因子以彩色椭圆表示。显示 HMG 蛋白（紫色）帮助弯曲 DNA，同时与增强体的其他蛋白相互作用。蛋白质大小和数量未按实际情况给出。

这种方式的典型特征是协同性，所有因子必须协同作用形成增强体。Mainatis 及同事证明 HMG I（Y）是协同性所必需的。在缺失 HMG I（Y）时，IFN-β 基因也可能被不同程度地激活，但当 HMG I（Y）存在时将大大促进其协同作用。这意味着当 HMG I（Y）存在时，只需很低浓度的其他激活因子即可最大限度地刺激基因转录。

增强体的三维结构是否在协同性中发挥作用？Mainatis 及同事注意到，所有的 HMG I（Y）结合位点都位于 DNA 螺旋的同一面。这种排列表明该蛋白一定结合在 DNA 的一边，以促使 DNA 弯曲形成协同作用所需的合适形状。为了阐明这一点，Mainatis 及同事将 6bp（约 1/2 圈螺旋）或 10bp（一个螺圈）的片段插入 IFN-β 增强子的 HMG I（Y）结合位点之间。他们发现，插入 6bp 片段时，HMG I（Y）结合位点的两个侧翼区置于 DNA 螺旋的两边，由此降低了协同作用，但插入 10bp 片段时，会保持结合位点在 DNA 螺旋的同一面，因此对协同性几乎无影响。

小结　激活因子 LEF-1 通过其 HMG 域与靶 DNA 的小沟结合并诱导 DNA 产生强烈弯曲。LEF-1 自身不能增强转录水平，但其所致的 DNA 弯曲可能有助于其他激活因子及通用转录因子的结合与相互作用，以刺激转录。另一种激活因子 HMG I（Y），可能在人 β-干扰素基因的调控区发挥相同的作用。这些蛋白质是结构转录因子，对于 IFN-β 增强子，激活作用需要包括 HMGI（Y）在内的几种激活因子协同结合，以形成一个特定形状的增强体。

隔离子

我们知道，增强子能对距离很远的启动子产生作用。例如，果蝇 *cut* 基因位点的翅缘增强子与启动子相隔约 85kb。在这么远的范围，某些增强子可能与其他不相关基因靠得足够近，产生激活作用。细胞怎样阻止这种不应有的激活呢？在高等生物中（至少包括果蝇和哺乳动物）是利用一种叫**隔离子**（insulator，或绝缘子）的 DNA 元件阻止附近的增强子对无关基因的激活作用。

Gary Felsenfeld 将隔离子定义为“相邻元件相互影响的中性阻碍物”。因此，大多数隔离子能保护基因免受附近增强子或沉默子（silencer）的激活或抑制作用。

隔离子具有两种功能被称为**增强子屏蔽**（enhancer blocking）和**阻碍物**（barrier）活性。增强子屏蔽是指隔离子具有阻止增强子激活转录效应的能力。阻碍物是指隔离子区域阻止染色质浓缩蔓延到靶基因区域的能力，从而防止基因发生沉默。并非所有隔离子都同时具有增强子屏蔽和阻碍物功能，某些隔离子被特化为只有其中一种功能。酵母中，对靠近着丝粒处的沉默子起阻碍作用的 DNA 元件就是一个只有阻碍物功能的隔离子的典型例子。

隔离子作用的机制目前还不十分清楚。但我们知道隔离子能够界定 DNA 结构域之间的边界。因此，在增强子和启动子之间加入隔离子可破坏原有的激活作用。同样，在沉默子和基因之间插入隔离子会消除抑制作用。所以，隔离子似乎是在基因区和增强子区（或沉默子区）形成一个边界，使基因不能感受到激活（或抑制）作用（图 12.28）。

我们也知道，隔离子的作用取决于所结合蛋白的特性。例如，某种果蝇隔离子包含 GAGA 序列，被称为 **GAGA 盒**（GAGA box），需 GAGA-结合蛋白 **Trl** 产生隔离子活性。遗传学实验证明，GAGA 盒本身发生突变，或 *trl* 基因（编码 Trl 蛋白）突变

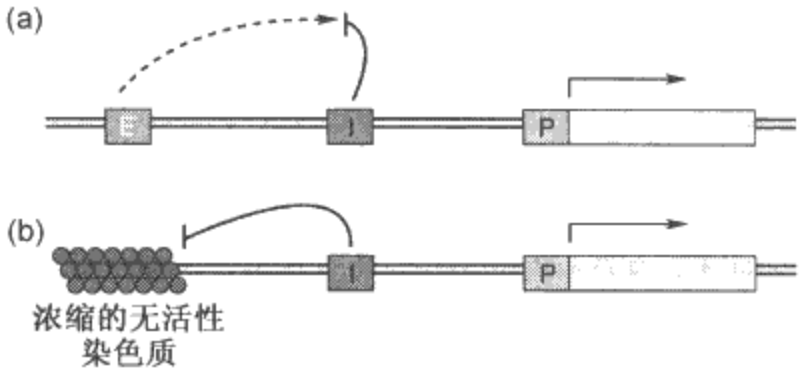


图 12.28　隔离子的功能。（a）增强子的屏障活性。在启动子和增强子之间的隔离子阻止启动子感受增强子的转录激活作用。（b）阻碍物活性。在启动子和浓缩、阻抑状态的染色质之间的隔离子，可阻止启动子受浓缩染色质（由沉默子诱导）对基因转录的阻抑的作用（实际上是阻止浓缩染色质侵蚀启动子）。

都能使隔离子丧失活性。

可以推测隔离子有多种作用机制，但最不可能的是隔离子诱导其上游染色质形成沉默的、浓缩的结构域。否则，位于隔离子上游的基因将总是被沉默了。但对果蝇的实验则表明这样的上游基因仍有潜在活性，并能被增强子激活。

图 12.29 给出了隔离子作用的另外两种模型。在第一种模型中，信号从增强子逐渐向启动子移动，而中间的隔离子阻止了该信号的前行。第二种模型需要位于增强子两端的隔离子间的相互作用，结果是增强子被隔绝在环形结构中，使其不能与启动子相互作用。

第一种假说很难解释 J. Krebs 和 Dunaway 的实验结果，该实验在理论上与前面介绍的 Dunaway 和 Dröge 的实验相近。在那个实验中（图 12.22）中，Dunaway 和 Dröge 将启动子和增强子分别置于连环体的两个不同 DNA 环上，并显示增强子仍有活性。后续实验中，Krebs 和 Dunaway 仍使用相同的连环体，但将两个果蝇的隔离子 *scs* 和 *scs′* 置于增强子或启动子两端。发现这两种情况中，隔离子都阻止增强子的活性。而单一隔离子在任一连环体质粒中对增强作用均无效。Dunaway 小组的两个实验是一致的，即只有当信号从一个 DNA 环跃至另一个环时，信号才能从增强子传向启动子。

也存在反对第二种假说的争议。其中主要的论点是，某些隔离子以单拷贝的形式起作用，并不一定要有两个隔离子分布在增强子的两侧。但有可能在这些实验中的第二个隔离子未被鉴定出来。第二个隔离子可吸引新蛋白与结合了已知隔离子的蛋白质相互作用，染色质因此而被环化，以阻止增强子与位于同一侧的启动子的相互作用。

Haini Cai 和 Ping Shen 的实验支持这一假说。

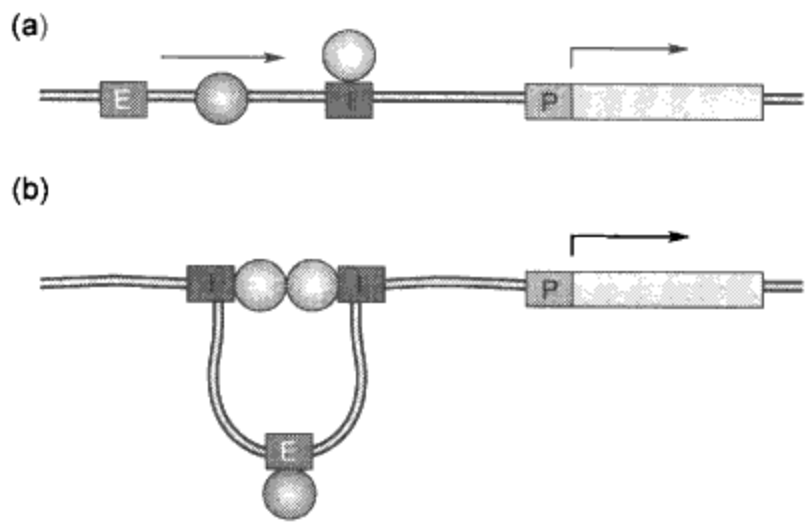


图 12.29　隔离子活性机制的两种假说。（a）滑动模型。激活因子结合到增强子上，刺激信号（绿色）或转录激活因子本身沿 DNA 滑向启动子。隔离子（红色）可能与一个或多个附着的蛋白质一起，阻挡滑向启动子的信号。（b）环化模型。两个隔离子（红色）位于增强子的两侧，当蛋白质（紫色）结合隔离子后，它们相互作用，将增强子隔离在环中，使其不能激活附近启动子（橙色）的转录。

他们在增强子和启动子之间插入一个已知的果蝇隔离子［su（Hw），毛翅基因抑制子］，观察到隔离子有一定的作用（增强效果减弱）。但是，将两拷贝的隔离子插入同一位置时，隔离子没有活性。最后，他们将两个 su（Hw）隔离子分别插入增强子的两侧，结果检测到最大的隔离子抑制效应。

图 12.30 说明了 Cai 和 Shen 对实验结果的解释。图 12.30a 表示单个隔离子的情况，它与一个未知隔离子（I）（在增强子上游的某处）连接起来抑制增强子的作用。图 12.30b 显示在增强子两端各有一个隔离子的情况。蛋白质与隔离子结合，促使 DNA 环化，使增强子被独立出来不再与启动子相互作用。图 12.30c 显示结合在增强子和启动子间的两个相邻隔离子上的蛋白质相互作用，使其间的 DNA 环化，但这不干扰增强子的作用，相反，由于使增强子更靠近启动子，转录增强的效果更显著。

在 2001 年的同一时间，Vincenzo Pirrotta 及合作者报道了他们利用单拷贝和多拷贝 su（HW）隔离子，对果蝇不同启动子开展的相同实验及相同结果。随后，他们增加了新内容，利用两个串联的不同基因，以及 3 个上游增强子及处于不同位置的 1～3 个隔离子，替代只有一个基因的情况。这 2 个基因分别是 *yellow* 和 *white*，其野生型使成年果蝇分别产生黑体色和翅色及红眼。当 *yellow* 失活（或突变）时，黑色素不能形成，后代身体和翅均为黄色而非黑色。当 *white* 基因失活（或突变）时，红眼色素不能合成，后代的眼睛为白色。

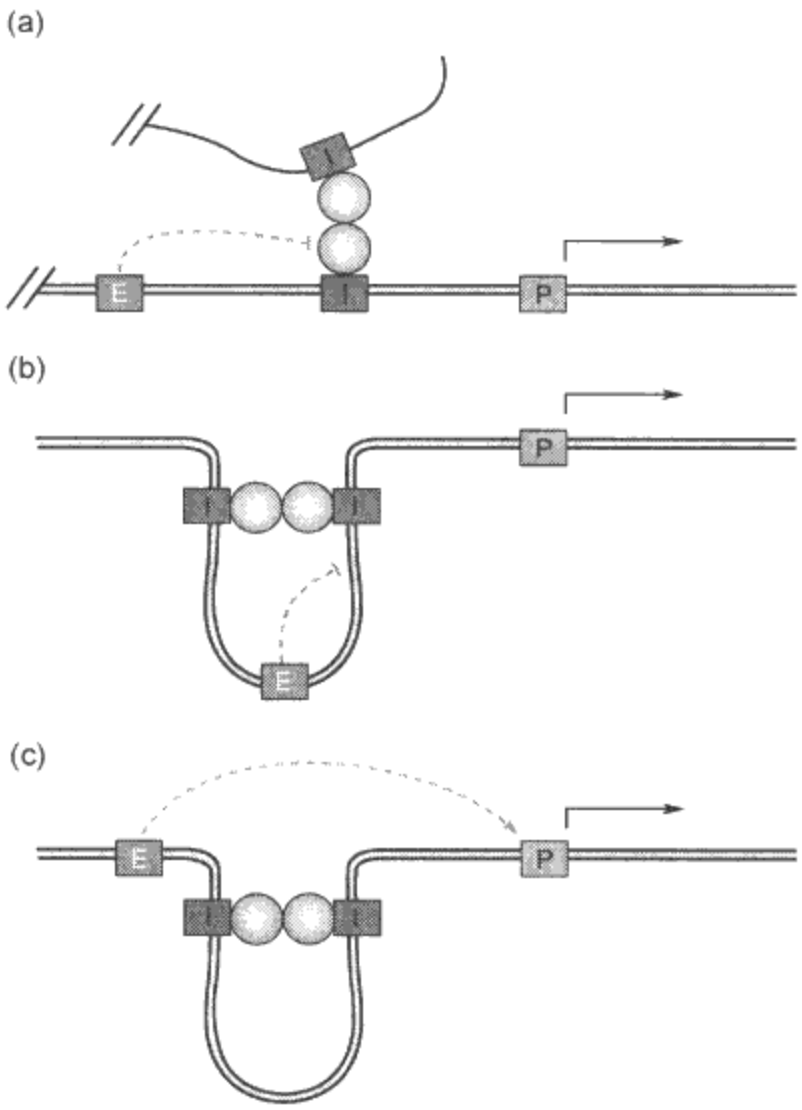


图 12.30　多隔离子的作用模型。（a）在增强子（E，蓝色）和启动子（P，橙色）之间的单一隔离子（I，红色）结合着蛋白质（紫色），这些蛋白质再与其他的结合同种或异种隔离子的蛋白质相互作用，使增强子与启动子隔开，阻止了转录增强作用。（b）增强子两侧的两个隔离子上所结合的蛋白质相互作用，使 DNA 环化，隔离增强子和启动子，阻止转录增强。（c）在增强子和启动子间的两个或多个隔离子上所结合的蛋白质之间相互作用使 DNA 环化，但未隔离增强子，反而使增强子更靠近启动子，结果使两个隔离子的作用相互抵消，不阻止转录的增强。这里，增强子和启动子可能通过 DNA 环化而靠近并相互作用，在此未详述。（*Source*：Adapted from Cai, H. N. and P. Shen, Effects of *cis* arrangement of chromatin insulators on enhancer-blocking activity. *Science* 291（2001）p. 495,（4.））

图 12.31 是 Pirrotta 及同事构建的质粒及实验结果。第 1 个质粒（EyeSYW）包含位于增强子和两个基因之间的单拷贝隔离子。正如 Cai 和 Shen 的模型所预测的，隔离子阻止增强子对两个基因的激活。第 2 个质粒（EyeYSW）的隔离子位于 *yellow* 和 *white* 基因之间。又如所预料，*yellow* 基因被激活，而 *white* 基因则没有。

第 3 个质粒（EyeSYSW）是两个隔离子分别位于 *yellow* 基因的两侧。结果更有意思，*yellow* 基因未被激活，而 *white* 基因被激活了。Cai 和 Shen 的模型再一次与这些结果相一致：*yellow* 基因两侧的隔离子阻止了该基因的激活，但却因此使两个隔离子同时位于增强子和 *white* 基因之间，隔离子的作用抵消，增强子激活 *white* 基因。因此，当两个隔离子的相互作用抵消了它们对 *white* 基因的作用时，并不真正地使隔离子失活，两个隔离子仍可阻止中间的 *yellow* 基因失活。第 4 个质粒（EyeSYWS）是两个隔离子分别位于 *yellow* 和 *white* 基因的两侧。果然，隔离子同时阻止了两个基因的激活。

最后，第 5 个质粒（EyeSFSYSW）含有 3 个隔离子，其中两个位于增强子和 *yellow* 基因之间，另一个在 *yellow* 和 *white* 基因之间。结果是两个基因都被激活。可见，增强子和基因间的两个至多个隔离子拷贝可中和隔离子的作用（在增强子和 *yellow* 之间有 2 拷贝，增强子和 *white* 之间有 3 拷贝）。我们可能会预期 *yellow* 基因上游的两个隔离子彼此作用相互中和，允许 *yellow* 基因激活，而 *yellow* 和 *white* 基因间的隔离子可能阻止 *white* 基因激活。而事实却是 3 个隔离子均失去了作用，两个基因均被激活。因此，该实验表明，两个串联隔离子的失活并非因为两者间简单的互斥作用造成。在某种程度上，结合于 3 个隔离子上的蛋白质之间的相互作用使上游增强子行使功能。

所有这些增强作用和隔离作用的结果都可以用图 12.30 的 DNA 成环模型很容易地解释。但 DNA 成环并非唯一解释。截至目前的实验证据还不能排除用某种追踪机制（图 12.20d）解释增强效应。在增强子和启动子间放一个隔离子，就能阻挡结合于增强子的蛋白质向启动子的追踪。那么如何解释在增强子和启动子间的两个或多个隔离子的抵消效应呢？一种方法是采用**隔离子小体**（insulator body，由两个或多个隔离子与其结合蛋白所形成的聚合体，这些蛋白质在细胞核周围已检测到）。尽管认为隔离子小体的形成在隔离子活性中起关键作用，但目前还没有关于隔离子小体作用机制的可接受假说。在这种情况下，不能排除这种可能，即两个或多个隔离子（增强子和启动子间）与其结合蛋白之间以某种方式相互作用，阻止了隔离子与隔离子小体的结合，进而阻止了隔离子活性。

或许总结我们对隔离子认识的最好方法就是承认隔离子的作用机制不止一种。有些隔离子以一种方式作用，而另一些隔离子以其他方式发挥作用。

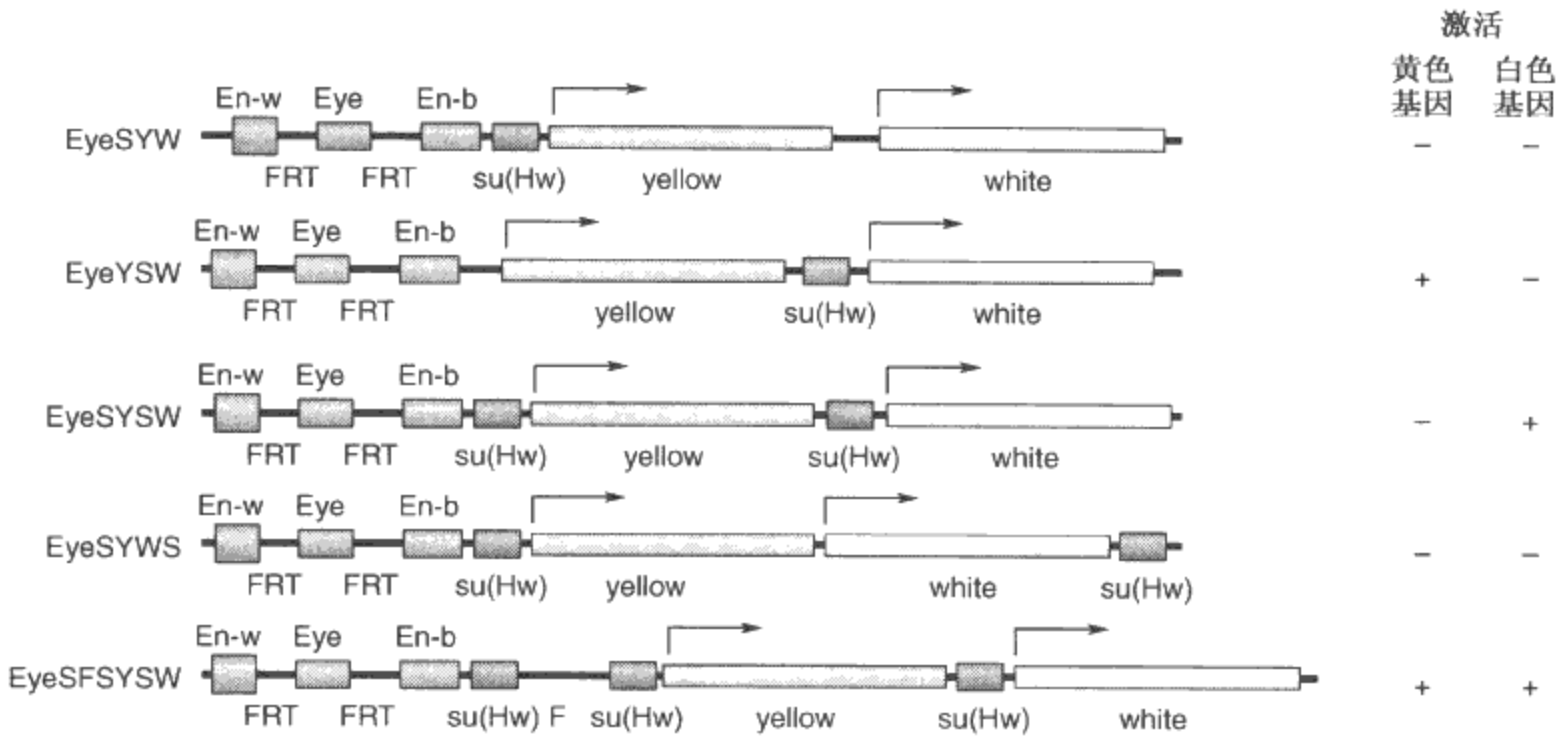


图 12.31　果蝇两个前后排列基因上的隔离子的作用。左边图示重组基因结构，结果［*yellow* 和 *white* 基因转录激活（＋）或者无转录激活（－）］标示在右边。重组基因的名字都以 Eye 开头，它表示 *yellow* 和 *white* 两个基因上游调控区的 3 个增强子（蓝色）中的眼睛特异性增强子。S、Y 和 W 分别代表隔离子［su（HW），红色］，*yellow* 基因和 *white* 基因。字母在重组基因名称中的位置表示其相应元件的位置。Pirrotta 等将重组基因转入果蝇胚胎中，并观察其体色和翅色（*yellow* 基因活性）和眼色（*white* 基因活性）的变化。（*Source*：Adapeted from Muravyova，E.，A. Golovnin，E. Gracheva，A. Parshikov，T. Belenkaya，V. Pirotta，and P. Georgiev，Loss of insulator activity by paired Su（Hu）Chromatin Insultators. *Science* 291（2001）p. 497，f. 2）

小结　隔离子是一种 DNA 元件，可屏蔽增强子对基因的激活作用（增强子屏蔽活性），或沉默子对基因的抑制作用（阻遏物活性）。有些隔离子兼具以上两种功能，有些则只有其中一种功能。隔离子可能通过成对合作发挥作用，两个隔离子上结合的蛋白质相互作用形成 DNA 环，这种环可隔离增强子和沉默子，使其不再激活或抑制启动子。这样，隔离子可能在染色体的不同 DNA 区域间形成边界。当两个或更多隔离子同时位于启动子和增强子之间时，彼此的作用相互抵消，可能是通过其上的结合蛋白相互作用，从而阻止 DNA 成环对增强子与启动子的隔离。另一种情况是，相邻隔离子结合蛋白间的相互作用阻止隔离子与隔离子小体的结合，这将阻止隔离子活性。

12.6　转录因子的调控

转录因子可正向或负向调控转录。那么这些转录因子又由谁来调控呢？本章前面已经遇到了这种例子，而在本章后面及第 13 章还会看到其他例子。现将其归类如下：

- 如本章前面学过的，核受体（如糖皮质激素受体）与其配体（如糖皮质激素）的结合可促使胞质内受体与抑制性蛋白解离，受体转移到细胞核并激活转录。
- 在第 13 章将看到，核受体与配体的结合可使核受体从转录阻遏物转变为激活因子。
- 激活因子的磷酸化（phosphorylation）使其与辅激活因子相互作用，激活转录。
- 转录因子泛素化（ubiquitylation）（使其附加泛素多肽）可将其标记进行蛋白质水解。
- 或者，转录因子泛素化激活其转录活性而不是降解。
- 转录因子的 SUMO 修饰（sumoylation，使其附加 SUMO 多肽）可将其导入细胞核的区室里，活性不能表达。
- 转录因子甲基化（methylation）能调节其活性。
- 转录激活因子乙酰化（acetylation）能调节其活性。

让我们来讨论一下其中的一些调控现象。

辅激活因子

某些 II 类激活因子也许自身就能召集基本转录复合物，这可能是通过结合一种或多种通用转录因

子或 RNA 聚合酶实现的，但多数转录因子却不行。Roger Kornberg 及同事首先提供了相关的证据。在 1989 和 1990 年，他们在研究**激活因子干扰**（activator interference）或**压制**（squelching）现象时，发现一定有其他因素介入其中。在体外转录实验中，当增加一个激活因子的浓度来抑制另一个激活因子活性时（也许是因为两种转录激活因子必须竞争某种稀有因子），就会发生压制现象。这种限制因子的合适候选者可能是某种通用转录因子，但加入大量通用转录因子并不能解除压制作用。这一事实表明，肯定存在另外某个因子为两种激活因子所必需。

那么，是什么因子呢？1990 年，Kornberg 及同事部分纯化了一种能解除激活因子压制作用的酵母蛋白。1991 年，他们进一步纯化了该因子并直接证明它有**辅激活因子**（coactivator）活性。即在体外，可刺激被激活的转录，但不是本底水平的转录。他们称其为**中介物**（mediator），因为它能介导激活因子的激活效应。利用酵母 *CYC1* 基因启动子和一个 GAL4 结合位点驱动的无 G 盒转录（第 5 章）进行转录分析，在激活因子 GAL4-VP16（嵌合激活因子，带有 GAL4 的 DNA 结合域及 VP16 的转录激活域）有或无条件下不断增加中介物的浓度。图 12.32 显示，在无激活因子条件下，中介物对转录没有影响（泳道 3～6）。有激活因子时，中介物对转录有极大促进作用（泳道 7～10）。用酵母激活因子 GCN4 所做类似实验的结果与此一致，显示中介物能与不止一个具有酸性激活域的激活因子协作。

类中介物复合物（mediator-like complex）也已从高等真核生物（包括人）中纯化出来。其中一个中介物由两个小组分别独立纯化出来，因此有两个不同的名字：**SRB 和含 MED 的辅因子**（SRB and MED-containing cofactor，**SMCC**），**甲状腺素受体相关蛋白**（thyroid-hormone-receptor-associated protein，**TRAP**）。SMCC/TRAP 是已知的哺乳动物类中介物复合物中最多的一种，但也有其他因子在结构和功能上与中介物相关联。其中之一是 CRSP，我们将在后面的内容中对它进行探讨。

在第 7 章我们学过，环化腺苷酸（cAMP）刺激细菌操纵子转录是通过结合激活因子 CAP，并使 CAP 与位于操纵子调控区的激活因子靶位点结合而实现的。cAMP 也参与真核生物的转录激活，但并不直接作用，而是通过一系列**信号转导途径**（signal transduction pathway）发挥作用。当真核细胞中 cAMP 水平升高时，cAMP 激活**蛋白激酶 A**（protein kinase A，**PKA**）的活性，并使该酶进入细胞核。在细胞核内，PKA 使 **cAMP 应答元件结合蛋白**（cAMP response element-binding protein，**CREB**）

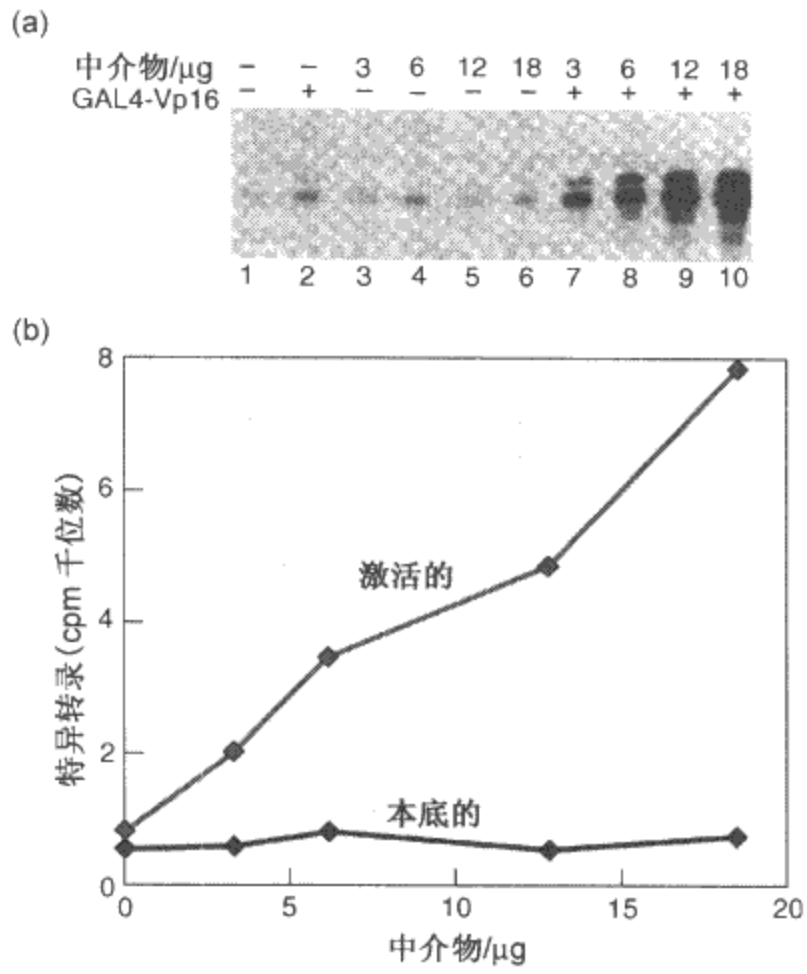


图 12.32　中介物的发现。Kornberg 及同事将酵母 *CYC1* 启动子置于 GAL4 结合位点的下游及无 G 盒的上游，因此无 G 盒的转录取决于 *CYC1* 启动子和 GAL4。然后在无 GTP 并加入不同浓度中介物［图（a）顶部标出］，在无（－）或有（＋）激活因子 GAL4-VP16 条件下，体外转录以上重组载体。用一种标记核苷酸标记体外转录物，电泳检测标记的 RNA。（a）凝胶电泳的磷酸成像仪扫描图像；（b）（a）中结果的曲线图。注意，有激活因子时，中介物极大地刺激转录，但对本底转录没有激活效果。（*Source*：Flanagan，P. M.，R. J. Kelleher，3rd，M. H. Sayre，H. Tschochner，and R. D. Kornberg，A mediator required for activation of RNA polymerase II transcription in vitro. *Nature* 350（4 Apr 1991）f. 2，p. 437. Copyright © Macmillan Magazines Ltd.）

磷酸化，随后 CREB 与 **cAMP 应答元件**（cAMP response element，**CRE**）结合，激活相关基因的转录。

因为 CREB 磷酸化是转录激活所必需的，因此预期这种磷酸化会使 CREB 进入细胞核或更牢固地结合 CRE。但事实上，这两种情况都不会发生。CREB 就定位在细胞核，甚至在未磷酸化的情况下就能很好地结合 CRE。那么，CREB 磷酸化如何使得转录的激活呢？答案的关键在于 1993 年发现的 **CREB 结合蛋白**（CREB-binding protein，**CBP**）。

CREB被蛋白激酶A磷酸化后，CBP能与CREB更好地结合。然后，CBP联系并召集基本转录装置的元件，或以整体形式召集聚合酶全酶。通过偶联CREB到转录装置上，CBP发挥辅激活因子作用（图12.33）。

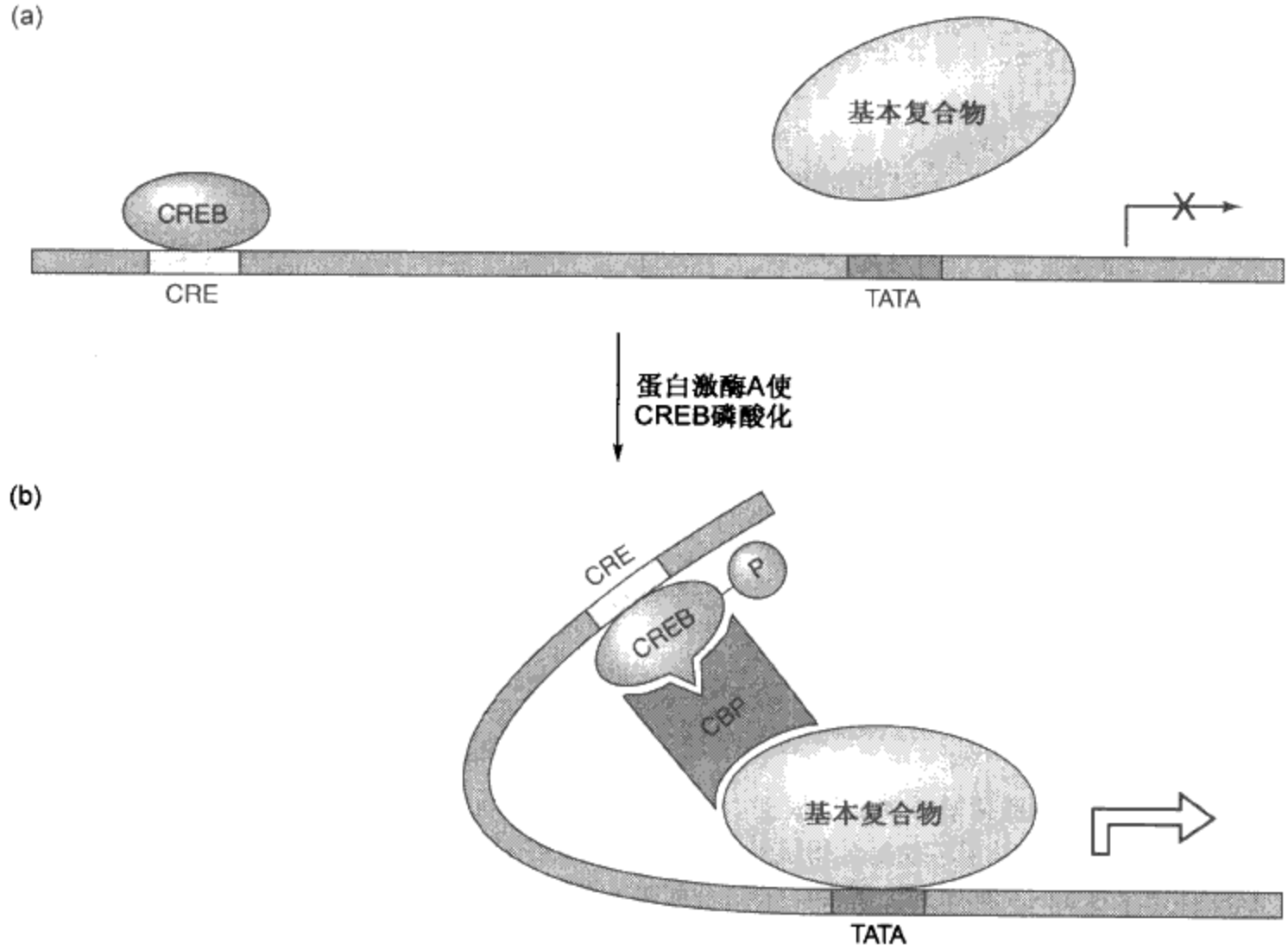


图12.33　CRE相连基因的激活模式。（a）未磷酸化的CREB（青绿色）与CRE结合，但是大量的基本复合物（RNA聚合酶与通用转录因子，橙色）不与启动子结合，甚至可能还未组装，基因不被激活；（b）PKA将CREB磷酸化，并使其结合CBP（红色）。随后，CBP与基本复合物中至少一个组分结合，召集基本复合物到启动子上，此时转录被激活。

自1993年发现CBP以来，已鉴定出许多辅激活因子。1999年，Tjian及同事分离出一种在体外转录中转录因子Sp1所需要的辅因子（cofactor required for Sp1 activation，CRSP）。纯化这种辅因子时，发现它有9个假定的亚基。用SDS-PAGE分离这些亚基，并转移至硝酸纤维素膜上，然后用蛋白酶水解各亚基，产生可测序的小肽段。测序结果显示，CRSP的有些亚基是独有的，但多数与已知的辅激活因子相同或至少同源，如酵母中介物的亚基。因此，不同辅激活因子可能是由其他各种辅激活因子的亚基重新混合组配而成。中介物和CRSP可能有共同作用模式，两者都与RNA聚合酶II的CTD结合。这种相互作用也许可以解释辅激活因子如何协助召集基本转录复合物。

CBP的辅激活因子作用不只限于cAMP应答基因，也适于核受体应答基因。这正好可以解释为什么不能检测到核受体转录激活域与任何通用转录因子间的直接作用。部分原因是核受体不直接接触基本转录装置，而是由CBP或其同源蛋白**p300**作为辅激活因子，协助将核受体和基本转录装置联系在一起。但CBP不能独立完成该任务，需另一家族的辅激活因子，**类固醇受体辅激活因子**（steroid receptor coactivator SRC family）合作进行。由于SRC的分子质量通常为160kDa，该类蛋白质有时也称为p160家族。SRC蛋白家族包括3个同源蛋白组**SRC-1**、**SRC-2**和**SRC-3**，它们与结合有配体的核受体相互作用。这种相互作用发生在核受体的激活域和SRC蛋白链中部的所谓LXXLL框（其中L代表亮氨酸，X代表其他任何氨基酸）之间。SRC蛋白也可与CBP结合，帮助核受体召集CBP，而CBP又进一步召集基本转录装置。首先发现的SRC家族成员是SRC-1（图12.34）。它与多种配体结合型受体相互作用，如孕激素受体（progestin receptor）、雌激素受体（estrogen receptor）、甲状腺激素受体（thyroid hormone receptor）。SRC-1不仅介导核受体与CBP的相互作用，还召集**CARM1**（**辅激活因子相关精氨酸甲基转移酶**，coactivator-associated arginine methyltransferase）。

CARM1 使启动子附近的蛋白甲基化。稍后将讨论　　CARM1 的功能。

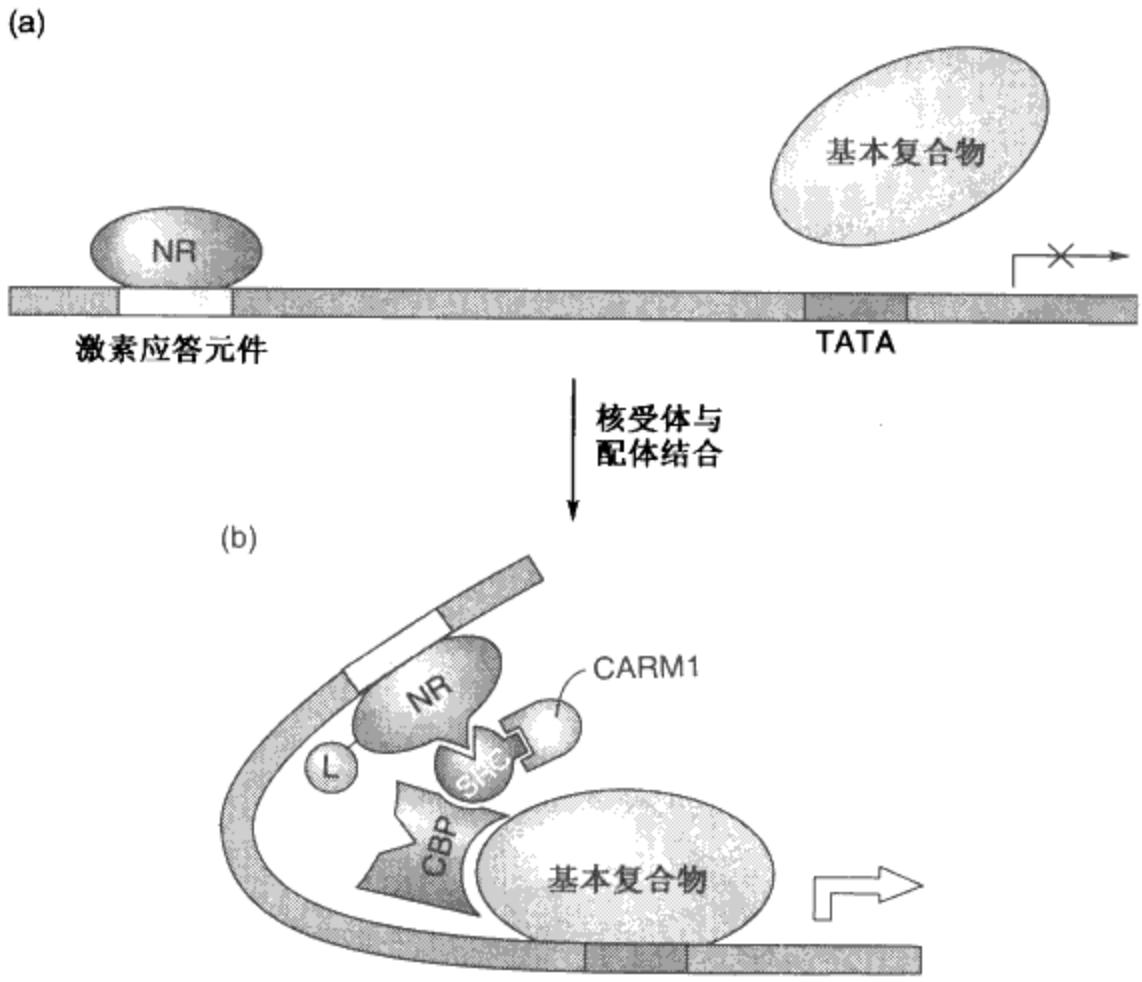


图 12.34　核受体激活基因的激活模型。（a）未结合配体的核受体（NR）结合在其激素应答元件上，但不能与基本转录复合物接触，所以相连基因不被激活。根据其类型，没有配体的核受体也能从 DNA 靶序列上解离。未结合配体的核受体在结合靶 DNA 后，也可有效地抑制转录。（b）核受体与配体（L,）结合，可以与 SRC 相互作用，SRC 又与 CBP 结合，而 CBP 又与基本转录装置的至少一个组分结合，并将转录装置召集到启动子上，激活转录。SRC 也与 CARM1 结合，它使启动子附近的蛋白质甲基化，进一步刺激转录。

还有一类重要的转录激活因子需要 CBP 作为辅激活因子。多种生长因子和细胞胁迫起始的级联事件（另一类信号转导途径），可导致磷酸化以及一种**分裂素激活蛋白激酶**（mitogen-activated protein kinase，**MAPK**）的激活。激活的 MAPK 进入细胞核并使激活因子如 Sap-1a 和 AP-1 中的 Jun 单体发生磷酸化。这些激活因子随后利用 CBP 介导靶基因的激活，最终刺激细胞的分裂。

除了召集基本转录装置到启动子上以外，CBP 还在基因激活方面发挥其他作用。CBP 具有很强的组蛋白乙酰转移酶活性，能将乙酰基附加在组蛋白上。在第 13 章将看到，组蛋白通常是基因活性的阻遏物，而组蛋白乙酰化可使其放松与 DNA 的结合，解除转录抑制作用。因此，激活因子与 CBP 在增强子上的结合使组蛋白乙酰转移酶靠近增强子，并使组蛋白乙酰化，从而激活邻近的基因。在第 13 章将详细讨论这一现象。

我们已经看到，CBP 和 p300 可作为大量激活因子（包括 CREB 和核受体）的辅激活因子。这表明 CREB 和核受体途径在利用相同的辅激活因子对不同基因的激活过程中，可能存在竞争作用。Ronald Evans 及同事发现，细胞限制这种竞争作用的一种方法是 CBP/p300 的甲基化。

核受体不仅吸引 CBP/p300，还有几个其他蛋白质。其中之一是 CARM1。CARM1 的甲基转移酶活性使组蛋白被 CBP/p300（第 13 章）乙酰化后，又在精氨酸处被甲基化，而此处的甲基化具有转录激活作用。但是，CARM1 也能使 CBP/p300 上的精氨酸甲基化。CBP/p300 的靶精氨酸位于召集 CREB 必需的所谓 KIX 域内，但对核受体-CBP/p300 之间的相互作用无影响。因此，CARM1 具有转录开关作用，通过阻断 CBP/p300（译者注：原文为 CBP/300）与 CREB 的相互作用，CARM1 可抑制 CREB 应答基因的转录，但通过甲基化邻近的组蛋白而激活核受体应答基因。

小结　几种不同转录激活因子，包括 CREB、核受体以及 AP-1，不是通过直接结合基本转录装置来激活转录，而是与 CBP（或其同源蛋白 p300）结合，CBP 再与基本转录装置作用并将其

召集到启动子上。与核受体响应元件结合的CBP/p300也能召集CARM1，CARM1使CBP/p300上与CREB相互作用的一个精氨酸甲基化，从而抑制CREB应答基因的激活。

激活因子的泛素化

有时，被激活的基因会因激活因子的降解而失活。例如，**LIM同源结构域**（LIM homeodomain，**LIM-HD**）家族的转录因子与辅阻遏物及辅激活因子结合。这些辅激活因子被称作**CLIM**，意为“LIM的辅因子”，辅阻遏物被称为**RLIM**，表示“环指LIM域-结合蛋白”（RING finger LIM domain-binding protein）。

CLIM蛋白与RLIM蛋白竞争结合LIM-HD激活因子。那么RLIM蛋白如何胜出，并抑制LIM-HD激活的基因呢？答案可能在于RLIM蛋白具有使LIM-HD结合的CLIM蛋白降解，并取而代之的能力。RLIM蛋白降解CLIM蛋白的方法是，先结合CLIM再对其赖氨酸残基附加多拷贝的**泛素**（ubiquitin）小分子蛋白质，形成所谓的**泛素化蛋白**（ubiquitylated protein）。一旦附加的泛素分子链足够长时，该蛋白质将被送入细胞质的**蛋白酶体**（proteasome）。蛋白酶体是个蛋白质集合，复合沉降系数为26S。蛋白酶体含有蛋白酶，可降解任何被送入蛋白酶体的泛素化蛋白。

与泛素相关的蛋白酶体行使着控制蛋白质质量的功能（quality control）。据估计，20%的细胞蛋白质由于转录或翻译错误而不正常。这些异常蛋白质对细胞具有潜在破坏性。因此，在它们造成危害之前就被泛素标记并送至蛋白酶体进行降解。其他正确合成的蛋白质也能被氧化或高温等胁迫所变性。为此，细胞有一种**伴侣蛋白**（chaperone protein）能够去折叠，使变性蛋白质重新正确折叠。有时变性很严重以致不能重新正确折叠，这种情况下，变性蛋白质将被泛素化，然后由蛋白酶体降解。

令人意外的是，泛素化对激活因子也有正效应。其中一个例子来自酵母的*MET*基因，它是含硫氨基酸（甲硫氨酸和半胱氨酸）合成所必需的。*MET*基因受甲酰供体浓度的调控，甲酰供体硫-腺苷甲硫氨酸，叫做SAM或AdoMet（S-adenosylmethionine，SAM or AdoMet）（第15章）。当SAM浓度低时，*MET*基因被激活因子Met4激活，当SAM浓度升高时，Met4被一个包括泛素化的过程所失活。这暗示Met4被泛素化并被蛋白酶体降解了，但情况并不是这么简单。

Met4降解确实在其失活中起作用，但在一定条件下（培养基富含甲硫氨酸），Met4即使被泛素化，也依然维持稳定状态。但这种状态不能激活*MET*基因，也不再与*MET*基因正确结合，尽管它仍能结合并激活另一类叫做*SAM*的基因。因此，Met4的泛素化虽不能使其降解，但能直接导致其失活。而且，这种失活是有选择性的。泛素化影响Met4对某些基因的激活，却并不影响Met4对另一些基因的激活作用。

几个研究显示，强转录因子倾向于被泛素化调节，接着被蛋白酶体降解。这样可快速关闭强激活因子对基因的高表达，为细胞对基因表达的调控提供了一定的灵活性。但同样，情况并不是简单的蛋白质降解。有些激活因子被单泛素化（monoubiquitylation）（附加单个泛素分子）所激活，而对该激活因子的多泛素化（poly ubiquitylation）却将其标记为降解。

最近，越来越多的证据显示蛋白酶体在转录调控中的另一种功能。已经发现蛋白酶体的19S调节颗粒中的蛋白质在活跃的启动子上与转录因子结合在一起。而且，**19S颗粒**（19S particle）可强烈刺激体外转录的延伸。还有，19S颗粒的部分蛋白质能够被激活因子GAL4召集到启动子上，包括ATP酶（在蛋白质降解之前的折叠过程所必需，但不参与蛋白质水解）。因此，19S颗粒蛋白的激活效应可能与蛋白质水解无关。Joan Conaway及同事推测这种蛋白酶体蛋白通过使转录因子至少部分去折叠而重塑模型以刺激转录的起始或延伸，或对两者均有刺激。

小结 RLIM蛋白（是LIM-HD的辅阻遏物）可结合LIM-HD的辅激活因子（如CLIM蛋白）并将其泛素化，结果使辅激活因子被26S蛋白酶体降解，由RLIM辅阻遏物代替CLIM的位置。某些激活因子泛素化（尤其是单一泛素化）具有激活效应，而多泛素化则使同一蛋白质标记为降解。蛋白酶体19S调节颗粒的蛋白质能通过重塑模型来激活转录因子从而刺激转录。

激活因子的SUMO修饰

SUMO修饰（sumoylation）是在目标蛋白的赖氨酸残基上附加1个或多个拷贝的101个氨基酸的多肽**SUMO**（small ubiquitin-related modifier，**泛素相关修饰因子**）。该过程的机制非常类似于泛素化，但结果却大相径庭。SUMO修饰的激活因子不被降解，而是被送入特定的细胞核区室中，保持稳定但不能与其靶基因作用。

例如，某些激活因子包括PML，即“早幼粒细胞白血病”（promyelocytic leukemia）通常就是SUMO修饰后被隐藏在叫做PML致癌基因域（POD）的核体中（nuclear body）。在早幼粒细胞白血病细胞中，POD受到干扰，释放的转录因子（包括

PML）到达并激活靶基因，从而导致白血病。

另一个实例与 Wnt 信号转导途径有关，此途径需要激活因子 β-catenin 进入细胞核并与结构转录因子 LEF-1（本章前面介绍过）协同激活特定基因的转录。LEF-1 可被 SUMO 修饰，隐藏于核体内。没有 LEF-1，β-catenin 无法激活其靶基因，Wnt 信号因此被阻断。我们知道，LEF-1 也可不依赖 Wnt 信号途径而激活其他基因，如 IFN-β。这种激活也可因 LEF-1 的 SUMO 修饰被阻断。另一种观点认为，LEF-1 也可与阻遏物（如 Groucho）搭档，而这种阻抑作用也可能因 LEF-1 的 SUMO 修饰而阻断。

小结 某些激活因子可被 SUMO 修饰（偶联一个 SUMO 小蛋白），使其隐藏于核体中，不能行使其转录激活功能。

激活因子的乙酰化

第 13 章我们将介绍碱性蛋白质组蛋白与 DNA 结合抑制转录。我们早已知道组蛋白的赖氨酸残基可被组蛋白乙酰转移酶（HAT）乙酰化，从而减弱组蛋白的抑制活性。最近，研究人员已经证明 HAT 可使非组蛋白激活因子和阻遏物乙酰化，这对乙酰化蛋白的活性会产生正向或负向调控。

肿瘤抑制剂蛋白 p53 就是乙酰化刺激转录活性的一种激活因子。辅激活因子 p300 具有 HAT 活性，使 p53 乙酰化，p53 活性的增加刺激了该激活因子靶基因的强烈转录。

p300 的 HAT 活性也会使抑制剂 BCL6 乙酰化。这种情况的结果与激活因子的激活有同样效果，即转录激活。

小结 非组蛋白激活因子和抑制剂能被 HAT 乙酰化，这种乙酰化作用具有激活或抑制基因转录的效应。

信号转导途径

在前几节讲到 CREB、Jun 和 β-catenin 的磷酸化都是信号转导途径（signal transduction pathway）的结果。因此信号转导途径对转录调控起主要作用。让我们来进一步探讨信号转导途径的概念并探讨一些例子。细胞外包裹着半透膜（细胞膜），使内容物不会流失，并保护细胞不受外来有害物质侵害。细胞内部和外围环境之间的这一屏障意味着在进化中必须获得一种机制使细胞能感受周围环境条件变化并做出应答。信号转导途径提供了这种机制。通常细胞对环境的应答需要基因表达的改变，而信号转导途径最终会导致转录因子的激活，进而激活一系列基因的表达。

图 12.35 概括了 3 种信号转导途径：蛋白激酶 A 途径、Ras-Raf 途径、核受体途径。前两种途径主要通过蛋白质磷酸化的级联放大激活各组分并最终激活转录。我们来详细探讨一下 Ras-Raf 途径，并了解该途径中异常组分如何使细胞生长失控并转为癌细胞的。

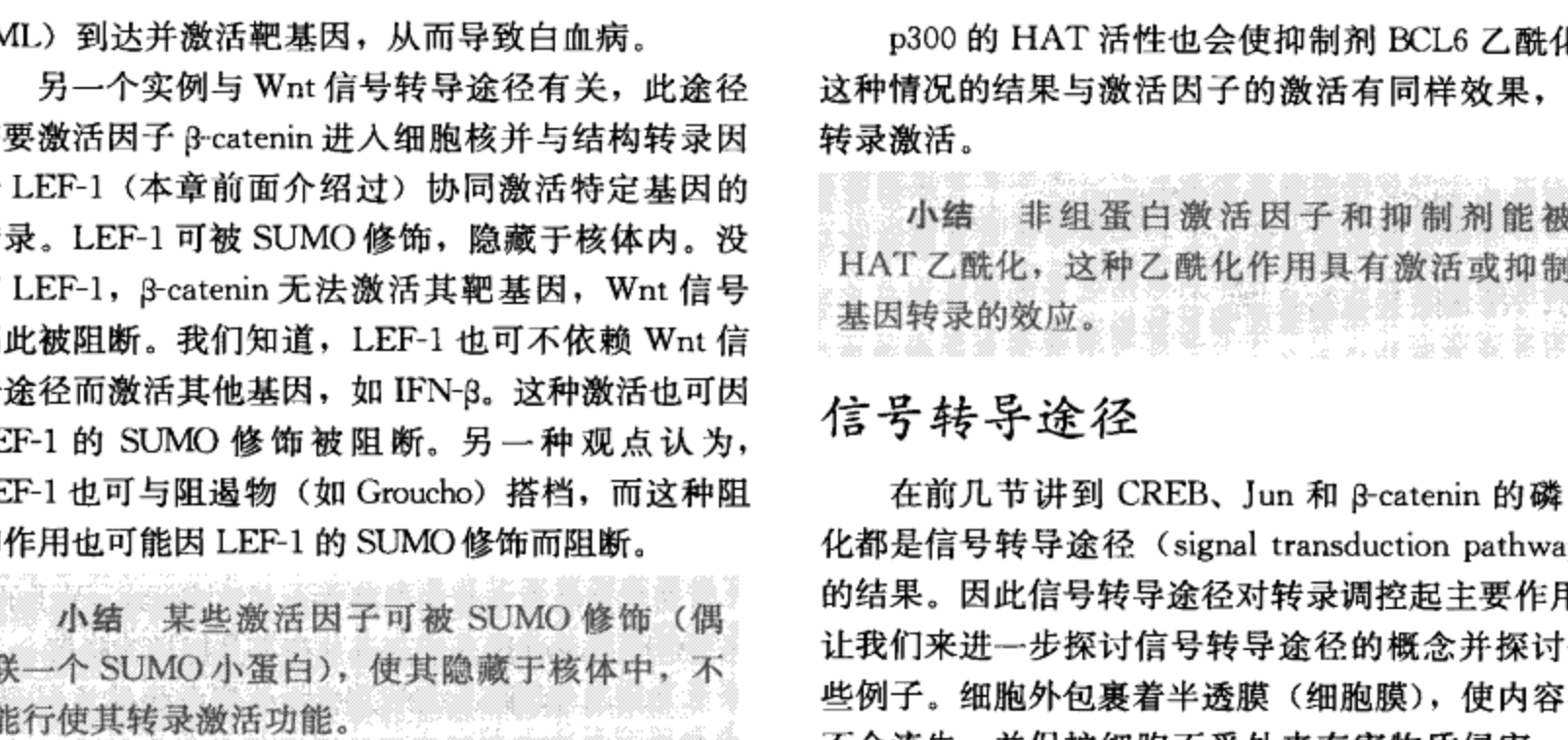


图 12.35 CBP/p300 的多种作用。3 种利用 CBP/p300 介导转录激活的信号转导途径均以 CBP/p300 为交汇中心。各组分的箭头仅表示在途经中所处的位置（如 MAPK 作用于 AP-1），并不说明反应的本质（如磷酸化）。简洁起见，该图还省略了各途经的分支。例如，MAPK 和 PKA 也使核受体磷酸化，尽管其重要性还不清楚。（*Source*：Adapted from Jankneht，R. and T. Hunter，Transcription：A growing co-activator network. *Nature* 383：23，1996. Copyright © 1996.）

图 12.36 显示 Ras-Raf 途径，其蛋白质是按照哺乳动物的系统进行命名的。在不同生物（如熟知的果蝇）中，同一途径的蛋白质有不同的命名。信号途径开始于胞外因子（如与细胞膜受体相互作用的生长因子）的诱导。当胞外因子，如**表皮生长因子**（epidermal growth factor，**EGF**）与其受体的胞外域结合后，刺激两个相邻受体靠近形成二聚体，并引起受体内具有酪氨酸激酶活性的胞内域彼此磷酸化。值得注意的是，跨膜受体如何将信号通过细胞膜传入细胞内（拉丁语，*transducere*，意为“穿过”）。一旦受体的胞内域被磷酸化，磷酸酪氨酸分子就会吸引具有特异性磷酸酪氨酸结合位点（称为**SH2 域**）的适配体（adapter）蛋白，如 **GRB2**（读作“grab two”）。SH2 域用以命名致癌蛋白 pp60^{src}（能将正常细胞转为类肿瘤细胞）上相似的位点，SH 表示“Src 同源”（Src homology）。GRB2 还有一个域，叫做 SH3（也在 pp60^{src} 中发现），它可吸引带有特殊疏水 α 螺旋的蛋白质，如 **Sos**。Sos 是一类 **Ras 交换因子**（Ras exchanger），能用 GTP 替换 **Ras** 上的 GDP，从而激活 Ras 蛋白。Ras 具有内在 GTP 酶活性，能将 GTP 水解为 GDP 使其自身失活。这种 GTP 酶自身活性很弱，但能被 **GTP 酶激活物蛋白（GTPase activator protein，GAP）** 强烈激活。因此，GAP 是此信号途径的抑制剂。

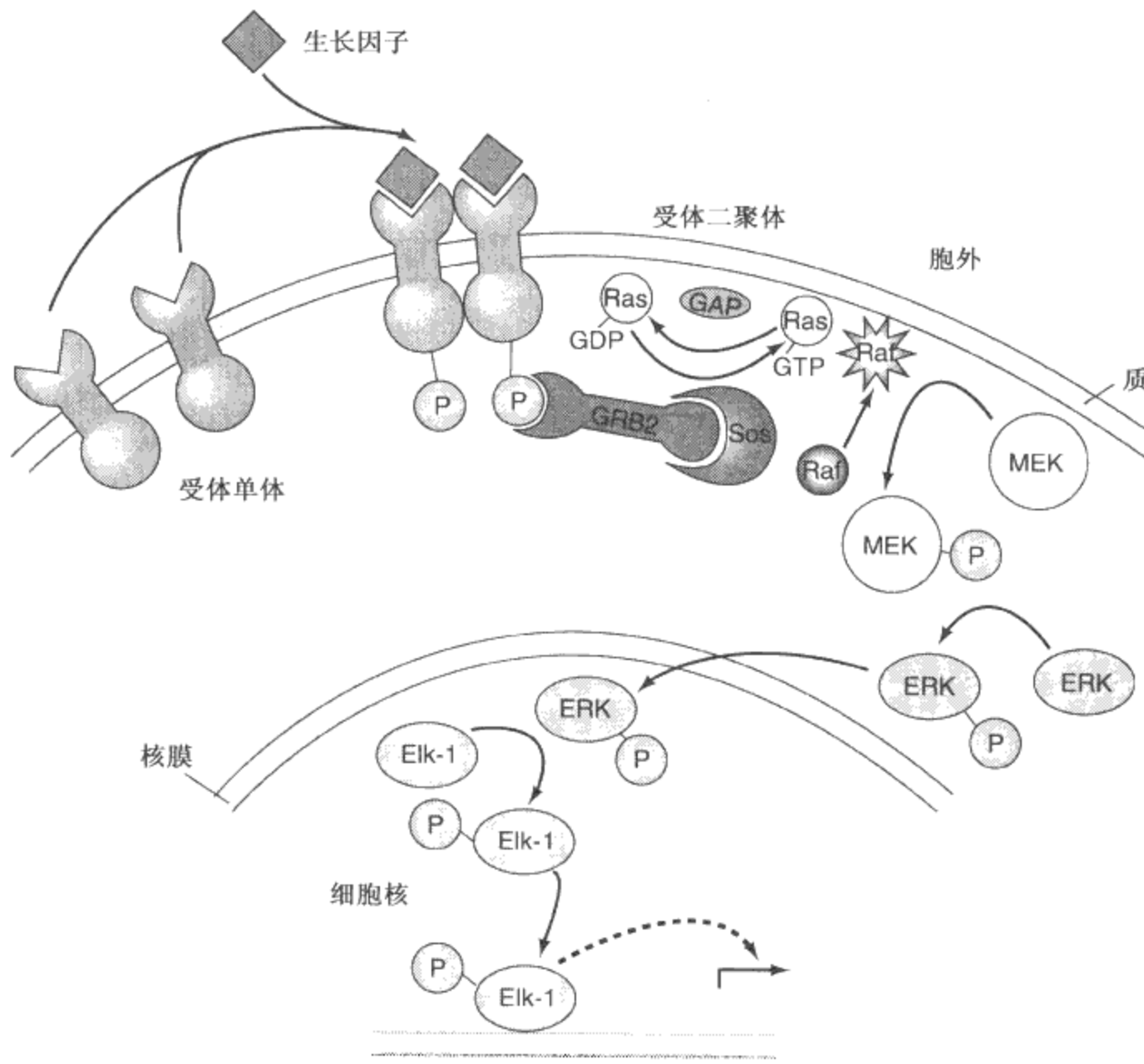


图 12.36　含有 Ras 和 Raf 的信号转导途径。当生长因子或其他胞外信号分子（红色）与受体（蓝色）结合后，信号转导就开始了（顶部）。受体结合配体后发生二聚化。每个受体单体的胞内蛋白酪氨酸激酶域彼此磷酸化。新的磷酸酪氨酸被适配体分子 GRB2（深绿色）识别，GRB2 再与 Ras 交换因子 Sos 结合。Sos（灰色）被激活后将 Ras 蛋白上的 GDP 置换为 GTP，从而激活 Ras（黄色）。Ras 将 Raf（紫色）送到细胞膜上变成激活态。Raf 的丝氨酸/苏氨酸激酶域在膜上被激活，进而使 MAPK/ERK 激酶（MEK，金色）磷酸化，激活的 MEK 又使胞外信号调节激酶（ERK，粉色）磷酸化。ERK 进入细胞核，将激活因子 Elk1（淡绿色）磷酸化。被激活的 Elk1 刺激了某些基因的转录，最终导致细胞分裂加速。

一旦途径被激活，Ras 将吸引另一个蛋白质 Raf 至细胞膜内表面，Raf 在这里被 Ras 激活。Raf 是另一种蛋白激酶，将磷酸基团加到丝氨酸和苏氨酸而不是酪氨酸上。Raf 的靶蛋白是另一种被称为 **MEK**

(**MAPK/ERK kinase，MAPK/ERK 激酶**）的丝氨酸激酶。随后，MEK 又使另一种蛋白激酶 **ERK**（**胞外细胞信号调控激酶**，extracellular-signal-regulated kinase）磷酸化并激活。激活的 ERK 可使大量胞质蛋白磷酸化，它还能进入细胞核使几种激活因子磷酸化并激活，包括 **Elk-1**。激活的 Elk-1 随后刺激那些促细胞分裂基因的转录。

至此，一个起始于生长因子与细胞表面相互作用，止于促生长基因转录增强的信号转导途径，可描述如下：

生长因子→受体→GRB2→Sos→Ras→Raf →MEK→ ERK→ Elk-1→转录增强→更多细胞分裂

该途径中编码许多组分的基因都是**原癌基因**（proto-oncogene）并不令人意外。原癌基因的突变将导致细胞生长失控及癌症。如果这些基因产生过多蛋白质产物，或其产物活性过高，整个途径就会加快，从而导致细胞生长异常加快，最终引发癌症。

值得注意的是该途径中信号的放大功能。一个 EGF 分子能激活若干个 Ras 分子，而每个 Ras 分子又激活若干 Raf 分子。由于 Raf 及随后的组分全部是激酶，每种酶能激活许多下一个分子。最终，一分子 EGF 将激活大量转录因子，从而引爆新的转录。还应注意，这仅是由一个 Ras 途径产生的效果。事实上，途径上有很多分支点，从而形成网状结构。不同信号途径组分间的这种相互作用叫做**信号交流**（cross talk）。

在本节最后介绍的 Wnt 信号转导途径（图 12.37）涉及我们讨论过的另外两种调控机制。首先，β-catenin 激活因子在 Wnt 信号不存在时被磷酸化。如同对其他激活因子一样，这种磷酸化不能激活 β-catenin，反而将其标记为泛素化，从而导致激活因子的降解。当 Wnt 信号存在时，β-catenin 磷酸化作用中止，激活因子不被泛素化，是稳定的，能够进入细胞核，并与 LEF-1 协同激活靶基因。

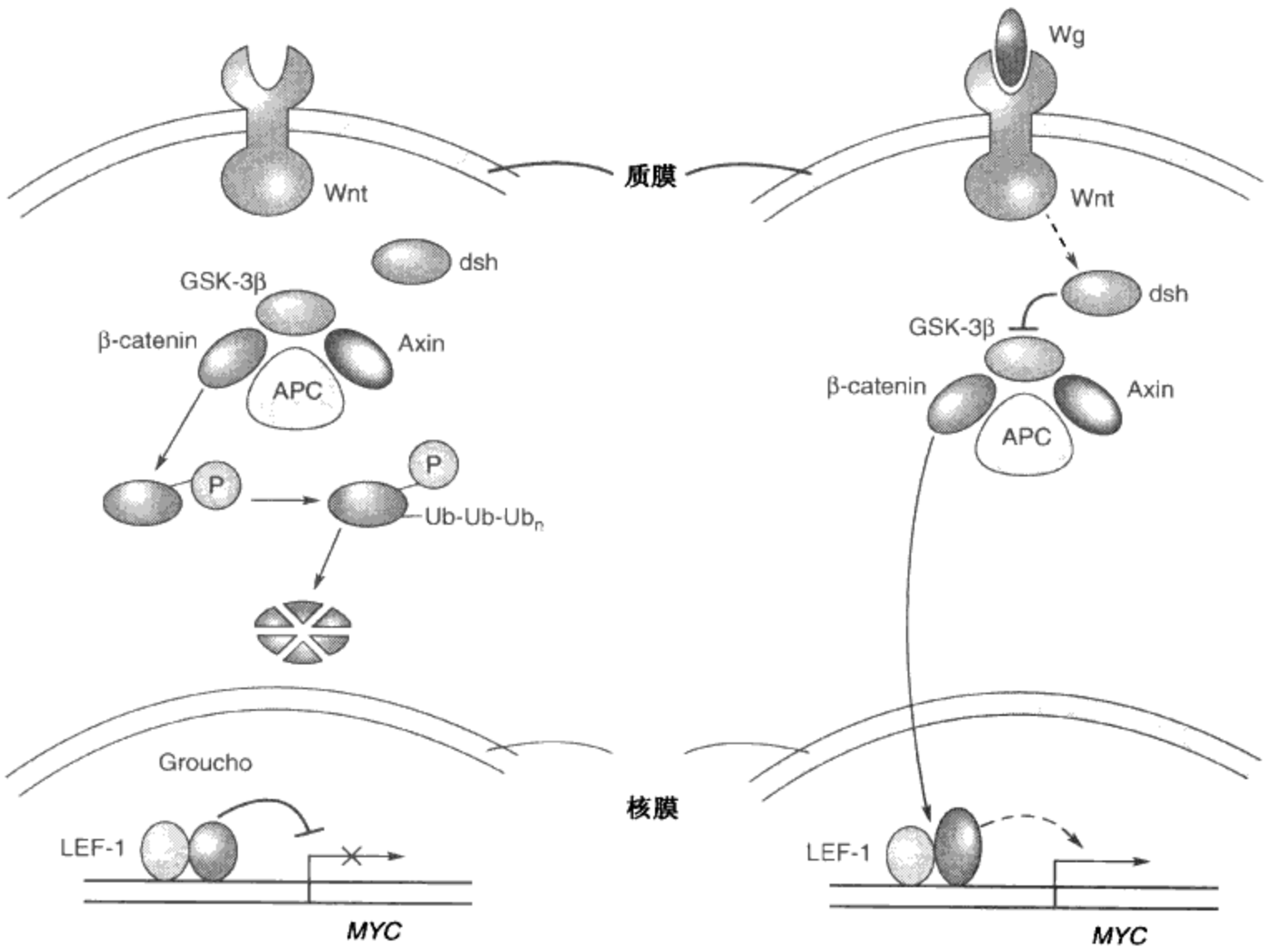


图 12.37　Wnt 信号途径。（a）无信号。无信号分子 Wg（Wingless）时，Wnt 受体不与 dsh（disheveled）发生作用。结果，蛋白激酶 GSK-3β［与至少 3 个其他蛋白质形成的复合体，包括 β-catenin（红色）］使 β-catenin 磷酸化。磷酸化的 β-catenin 被泛素化，进而被 26S 蛋白酶体降解。因此，β-catenin 不能激活任何靶基因，包括原癌基因 *MYC*。LEF-1 与抑制剂 Groucho 结合，协助抑制 *MYC*。（b）有 Wnt 信号。Wg 结合 Wnt，从而将信号传给 disheveled（dsh），该蛋白质以某种方式阻断 GSK-3β 的蛋白激酶活性（对 β-catenin 磷酸化）。由于 β-catenin 未被磷酸化，因此不会因泛素化而降解。最后，β-catenin 进入细胞核并协助 LEF-1 共同激活 *MYC* 的转录。

小结 信号转导途径通常始于一个与细胞表面受体相互作用的信号分子，并由受体将信号送达胞内，从而导致基因表达的改变。许多信号转导途径，包括 Ras-Raf 途径，依赖蛋白质磷酸化在蛋白质间传递信号分子，并使信号在每一步得到放大。

总结

真核细胞的转录激活因子至少有两个域：DNA 结合域和转录激活域。DNA 结合域包含像含锌组件、(部分) 同源结构域、亮氨酸拉链和螺旋-环-螺旋等模体。转录激活域有酸性域、富谷氨酰胺域和富脯氨酸域。

锌指由一个反向平行的 β 折叠及紧邻的一个 α 螺旋组成。β 折叠中包含一对半胱氨酸，α 螺旋包含一对组氨酸，它们与锌离子螯合。氨基酸与金属离子的螯合有助于形成指结构。指结构与其 DNA 靶序列的特异性识别发生于 DNA 双螺旋大沟内。

GAL4 蛋白的 DNA 结合模体含有 6 个半胱氨酸及 2 个锌离子形成双金属巯基簇。DNA 识别组件的短 α 螺旋插入 DNA 双螺旋大沟并在那里进行特异性相互作用。GAL4 单体也包含一个 α 螺旋二聚化模体，能在与其他 GAL4 单体相互作用时形成平行的螺旋卷曲。

I 型核受体与另一种蛋白质结合并位于细胞质中。当外源激素配体与 I 型受体结合后，释放偶联的胞质蛋白，并进入核内与增强子结合，行使激活因子的功能。糖皮质激素受体是该类群的代表。它的 DNA 结合域有 2 个含锌组件。一个组件含有大部分与 DNA 结合相关的氨基酸（在 α 识别螺旋中），另一组件则提供了蛋白质-蛋白质相互作用形成二聚体的触点。这些含锌组件利用 4 个半胱氨酸与锌离子配位结合，这与典型的锌指结构中两个半胱氨酸和两个组氨酸与锌离子配位不同。

真核细胞转录激活因子的同源结构域包含的 DNA 结合模体类似原核细胞的螺旋-转角-螺旋模体的作用原理，由一个识别螺旋嵌入 DNA 大沟并进行特异性相互作用。

bZIP 蛋白通过亮氨酸拉链而二聚化，使每个单体的相邻碱性区域像一把钳子一样夹住 DNA 靶序列。类似地，bHLH 蛋白通过螺旋-环-螺旋模体而二聚化，使每个长螺旋的碱性部分抓住 DNA 靶序列，很像 bZIP 蛋白那样。bHLH 和 bHLH-ZIP 域都以相同方式结合 DNA 靶序列，但后者由于其亮氨酸拉链而有额外的二聚化潜力。

激活因子蛋白的 DNA 结合域和转录激活域是相互独立的组件。杂合蛋白质兼具一种蛋白质的 DNA 结合域和另一种蛋白质的转录激活域，仍有激活因子功能。

激活因子通过与通用转录因子接触，刺激前起始复合物在启动子上的组装。体外观察发现，在 II 类启动子上，前起始复合物的组装是由通用转录因子和 RNA 聚合酶 II 按一定顺序完成的，或者是通过召集包括 RNA 聚合酶和大多数通用转录因子在内的聚合酶全酶来完成。某些因子（也许就是 TBP 或 TFIID）需独立于全酶之外被召集。

二聚化的巨大优势在于其增加激活因子和 DNA 的亲和力。有些激活因子形成同源二聚体，有些则形成更高效的异源二聚体，如 Jun 和 Fos。

增强子作用的本质是结合在增强子上的激活因子与结合在启动子上的通用转录因子及 RNA 聚合酶之间发生的相互作用。这可能以增强子和启动子之间 DNA 的环凸介导，至少在理论上是这样。DNA 环凸可以将结合在不同增强子上的激活因子拉至足够靠近启动子的位置，然后协同激活基因转录。

复合增强子能使一个特定基因对各种激活因子的不同组合产生应答。这种排布格局使细胞能够对基因在不同组织或不同发育阶段产生精细调控。

激活因子 LEF-1 通过其 HMG 域与靶 DNA 的小沟结合并诱导 DNA 产生强烈弯曲。LEF-1 自身不能增强转录水平，但其所致的 DNA 弯曲可能有助于其他激活因子与通用转录因子的结合和相互作用，从而刺激转录。另一种激活因子 HMG I（Y），可能在人 IFN-β 基因的调控区发挥相同的作用。这些蛋白质是结构转录因子，对于 IFN-β 增强子，激活作用需要包括 HMGI（Y）在内的几种激活因子协同结合，以形成一个特定形状的增强体。

隔离子是一种 DNA 元件，可屏蔽增强子对基因的激活作用（增强子屏蔽活性），或沉默子对基因的抑制作用（阻遏物活性）。有些隔离子兼具以上两种功能，有些则只具其中一种功能。隔离子可能通过成对合作发挥作用，两个隔离子上结合的蛋白相互作用形成 DNA 环，这种环可隔离增强子和沉默子，使其不能再激活或抑制启动子。这样，隔离子可能在染色体的不同 DNA 区域间形成边界。当两个或更多隔离子同时位于启动子和增强子之间时，彼此的作用相互抵消，可能是通过其上的结合蛋白相互作用，从而阻止 DNA 成环对增强子与启动子的隔离。另一种情况是，相邻隔离子结合蛋白间的相互作用阻止隔离子与隔离子小体的结合，这可阻止隔离子活性。

几种不同转录激活因子，包括 CREB、核受体以及 AP-1，不是通过直接结合基本转录装置来激活转录，而是与 CBP（或其同源蛋白 p300）结合，CBP 再与基本转录装置作用并将其召集到启动子上。与核受体响应元件结合的 CBP/p300 也能召集 CARM1，

CARM1 使 CBP/p300 上与 CREB 相互作用的一个精氨酸甲基化，从而抑制 CREB 应答基因的激活。

某些激活因子和辅激活因子受泛素介导的降解作用所调控。这些蛋白质被泛素标记后，由 26S 蛋白酶体进行降解。一些激活因子的单泛素化会产生激活效应，而多泛素化则使同一个蛋白质标记为降解。来自蛋白酶体的 19S 调节颗粒的蛋白质能通过重构模型来激活转录因子从而刺激转录。

某些激活因子可以被 SUMO 修饰（偶联一个类泛素小蛋白 SUMO），使它们被隔离在核体内，不能行使转录激活功能。非组蛋白激活因子和抑制剂能被 HAT 乙酰化，从而产生激活或抑制效应。

信号转导途径通常始于一个与细胞表面受体相互作用的信号分子，并由受体将信号送达胞内导致基因表达改变。许多信号转导途径，包括 Ras-Raf 途径，依赖蛋白质磷酸化在蛋白质间传递信号分子，并使信号在每一步得到放大。激活因子和信号转导途径上的其他组分的泛素化和 SUMO 修饰也能在这些途径中产生重要影响。

复习题

1. 列举真核激活因子 3 类不同的 DNA 结合域。

2. 列举真核激活因子 3 类不同的转录激活域。

3. 图示锌指结构，指出指形结构中的 DNA 结合模体。

4. 列举典型原核生物螺旋-转角-螺旋域和 Zif268 锌指域的一个重要相同点和 3 个不同点。

5. 图示与 DNA 相互作用的两个 GAL4 蛋白 N 端 65 个氨基酸形成的二聚体结构图。图中显示二聚体域和两个 DNA 结合域中与 DNA 结合位点相互作用的模体，显示每个 DNA 结合域中的金属离子及配位氨基酸（包括其数目）。

6. 在一般意义上，核受体的功能是什么？

7. 解释 I 型和 II 型核受体的区别，每种举出一个实例。

8. 列出核受体的每个 DNA 结合域中的金属离子及配位氨基酸（包括其数目）。DNA 结合域的哪一部分与 DNA 的碱基相互作用？

9. 同源结构域的本质是什么？它与哪种 DNA 结合域最相似？

10. 图示亮氨酸拉链底面观，并由图解释亮氨酸拉链结构和功能间的关系。

11. 图示 bZIP 蛋白与 DNA 结合位点相互作用的示意图。

12. 描述并解释基因特异转录因子的 DNA 结合域与转录激活域彼此独立的实验结果。

13. 举出召集 II 类前起始复合物的两种模型，其中一个涉及聚合酶全酶。

14. 叙述实验并给出结果，显示酸性转录激活域结合 TFIID。

15. 举出有利于全酶召集模型的证据。

16. 举出两个质疑全酶召集模型的证据。

17. 为什么蛋白二聚体（或四聚体）在 DNA 结合中比单体更有效？为什么与 DNA 特异性序列的高亲和力对转录激活因子很重要？

18. 列举 3 种模式解释增强子如何作用于相距几百个碱基以外的启动子。

19. 叙述实验并给出结果，显示连环体 DNA 一个环含增强子另一个环含启动子，对两个 DNA 元件的分离效果。该实验倾向于哪种增强子活性模型？为什么？

20. 复合增强子有什么优势？

21. 解释 LiCl 如何使海胆胚胎发育过程中内胚层过度生长。

22. LEF-1 是人 T 细胞受体 α-链的激活因子，但 LEF-1 自身无法激活该基因。LEF-1 如何发挥作用？给出支持答案的证据。

23. LEF-1 与 DNA 靶序列的大沟还是小沟结合？给出支持答案的证据。

24. 隔离子有什么作用？

25. 绘制模型解释如下的结果：（a）在增强子和启动子之间的一个隔离子部分抑制增强子活性；（b）在增强子和启动子之间的两个隔离子不抑制增强子活性；（c）在增强子任意一边的一个隔离子会严重抑制增强子活性。

26. 在增强子和启动子间有 3 拷贝的隔离子会出现什么效应？如何解释该现象？

27. 描述验证中介物效应的实验及结果。

28. 图示 CBP 作为辅激活因子的作用：（a）磷酸化的 CREB；（b）核受体。

29. 信号转导途径如何放大信号？举例说明。

30. 给出并解释泛素对转录的负调控假说。

31. 给出并解释蛋白酶体对转录的正调控假说。

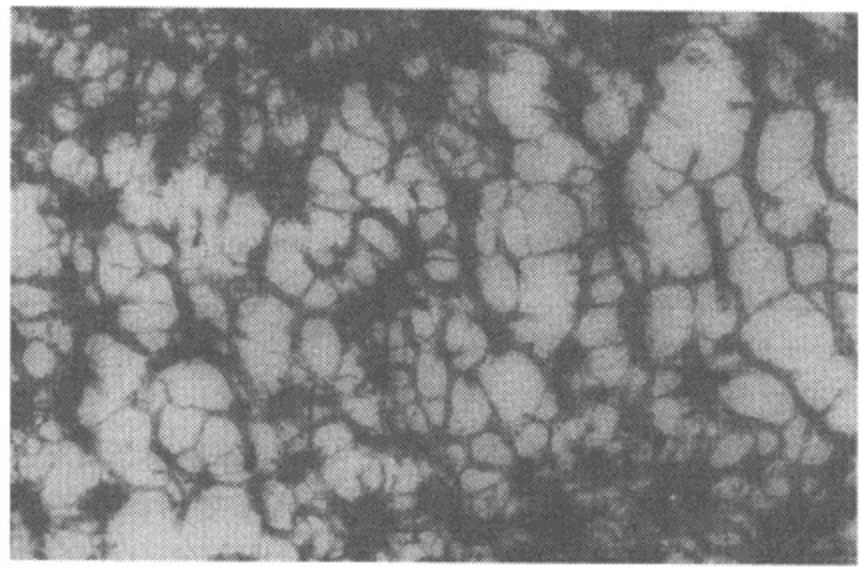
分析题

1. 设计一个实验，显示 TFIIB 直接与酸性激活域结合，并显示正向激活结果。

2. 假设你在研究由 3 个增强子调控的人 *BLU* 基因。推测可能结合在 3 个增强子上的蛋白质间相互作用形成激活所需的增强体。这种相互作用需要增强子间具有怎样的间隔序列？在间隔序列中引入怎样的改变来验证你的假说？期望得到什么结果？

翻译　兰海燕　校译　马 纪

第 13 章　染色质结构及其对基因转录的影响



人类发育中的精细胞染色质（×300 000）*Copyright © David M. Phillips/Visuals Unlimited.*

在讨论真核基因转录时，我们一直在忽略一个重要的事实，即天然真核基因并不以裸 DNA 分子状态存在，甚至也没有仅结合了转录因子的 DNA 分子。相反，它们与其他蛋白质，特别是一类称为组蛋白的碱性蛋白质形成复合物，这种复合物被称为**染色质**（chromatin）。我们将看到，组蛋白能抑制基因转录，而多年来一直认为组蛋白是一类毫无意义的基因转录阻碍物。现在我们意识到组蛋白与 DNA 和转录因子以动态方式相互作用调控基因的表达。本章先介绍组蛋白和染色质的结构，然后探讨组蛋白与其他染色质蛋白合作对转录进行调控的方式。

13.1　组蛋白

大多数真核细胞含有 5 类不同的组蛋白（histone）：H1、H2A、H2B、H3、H4。图 13.1 显示这 5 种组蛋白可以通过凝胶电泳分离。组蛋白是含量极丰富的蛋白质，在真核细胞核中组蛋白的量与 DNA 的量是相同的。组蛋白也是碱性极强的蛋白质，其中至少 20%的氨基酸是精氨酸或赖氨酸，在中性 pH 环境中带较强的正电荷。根据这个特性，可用强酸，如 1.5mol/L HCl，从细胞中提取组蛋白，而这种条件足以破坏大多数蛋白质。也正是由于这种碱性的本质，组蛋白在非变性电泳中向阴极泳动，而大多数蛋白质呈酸性，电泳中向阳极泳动。大多数组蛋白在生物间具有高度保守性，组蛋白 H4 就是极端的例子。牛组蛋白 H4 与豌豆 H4 相比，在总共 102 个氨基酸中仅有 2 个氨基酸不同，且都是保守性的改变，即一种碱性氨基酸（赖氨酸）替代另一种碱性氨基酸（精氨酸），以及一种疏水性氨基酸（缬氨酸）替代另一种疏水性氨基酸（异亮氨酸）。换句话说，起源于同一祖先的两个物种——牛和豌豆在漫长的历史进化中，它们的组蛋白 H4 只有 2 个氨基酸发生了改变。组蛋白 H3 在进化上也极端保守，组蛋白 H2A 和 H2B 中度保守，但组蛋白 H1 在不同物种间变异较大。表 13.1 列举了组蛋白的某些特性。

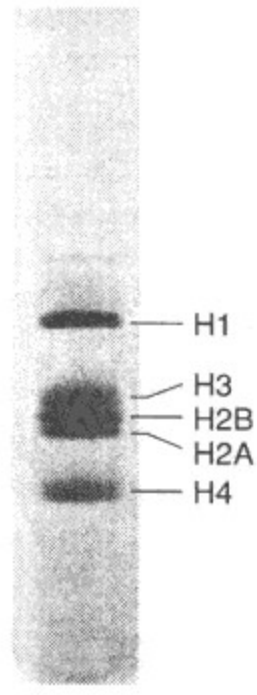


图 13.1　电泳分离组蛋白。聚丙烯酰胺凝胶电泳分离小牛胸腺组蛋白。（*Source*: Panyim and Chalkley. *Archives of Biochem. & Biophys.* 130, 1969, f. 6A, p. 343. Reprinted with permission of Academic Press.）

表 13.1　组蛋白的基本性质

组蛋白类型	组蛋白	相对分子质量（M_r）	保守程度
核心组蛋白——能与 DNA 组装成颗粒	H3	154 000	高度保守
	H4	11 340	高度保守
核心组蛋白——位于核心颗粒的外围	H2A	14 000	组织与物种间中度变异
	H2B	13 770	组织与物种间中度变异
类 H1 组蛋白	H1	21 500	组织与物种间显著变异
	H1°	约 21 500	易变，多数出现在非复制细胞
	H5	21 500	非常易变，仅出现于某些物种的无转录活性细胞

Source: © 1983, From Critical Reviews in Biochemistry and Molecular Biology, by Butler, P. J. C. Reproduced by permission of Taylor & Francis Group. LLC., http://www.taylorandfrancis.com

图 13.1 给人一种组蛋白一个类型的印象。然而，通过高分辨率分离，组蛋白表现出很大的多样性。造成这种多样性有两个原因，基因重复和翻译后的修饰。在真核生物中大多数蛋白质编码基因是单拷贝的，而组蛋白基因则呈高拷贝：小鼠为10～20 拷贝，果蝇约为 100 拷贝。其中的许多拷贝是相同的，但有些则完全不同。组蛋白 H1（富含赖氨酸组蛋白）变异最大，在小鼠中至少有 6 个亚型。有一种 H1 变种被称为 H1°。在鱼类、鸟类、两栖类和爬行类等动物中有另一种富含赖氨酸的组蛋白，它可能是 H1 的一种极端变种，但是它与 H1 的差异太大，通常又将其称为 H5。组蛋白 H4 的变异最小，迄今只报道了两个变种且极为少见。人们推测特定组蛋白的各个变种在本质上具有相同的功能，但每一种都对染色质特性有不同程度的影响。

组蛋白异质性的第二个原因是翻译后修饰，它是变异的主要来源，以至于 Kensal van Holder 把它称为“baroque”（意为结构复杂怪异的）。最常见的组蛋白修饰方式是乙酰化，发生在 N 端氨基基团和赖氨酸 ε-氨基基团上。其他修饰方式还有赖氨酸 ε-氨基的甲基化和磷酸化、丝氨酸和苏氨酸 *O*-磷酸化以及赖氨酸和组氨酸 *N*-磷酸化。这些修饰是动态过程，所修饰的基团能被除去，也可添加。至少某些组蛋白的修饰会影响染色质的结构和功能，并且在调控基因活性方面发挥重要作用。我们将在本章的后面讨论这一现象。

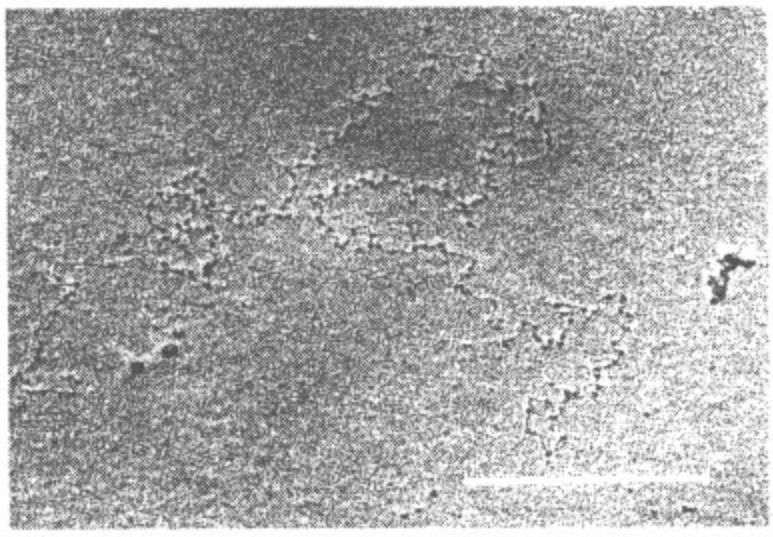
13.2　核小体

典型人类染色体的长宽比超过 1000 万∶1，如此长而细的分子如果不能有序地折叠起来将会纠结在一起造成混乱。此外，如果将人类一个细胞中的全部 DNA 连接起来，将达到 2m 长，而这样长的 DNA 必须被包裹在直径仅有 10μm 的细胞核内，因此必须考虑如何折叠的问题。事实上，如果把人体内所有 DNA 分子首尾相接，其长度足以从地球到太阳来回很多次。因此，在人类或其他生物体中就必然发生 DNA 的高度折叠。我们将看到真核细胞 DNA 确实以多种方式进行折叠。一级折叠形成的结构称为**核小体**（nucleosome），它有一个组蛋白的核，外面缠绕着 DNA。

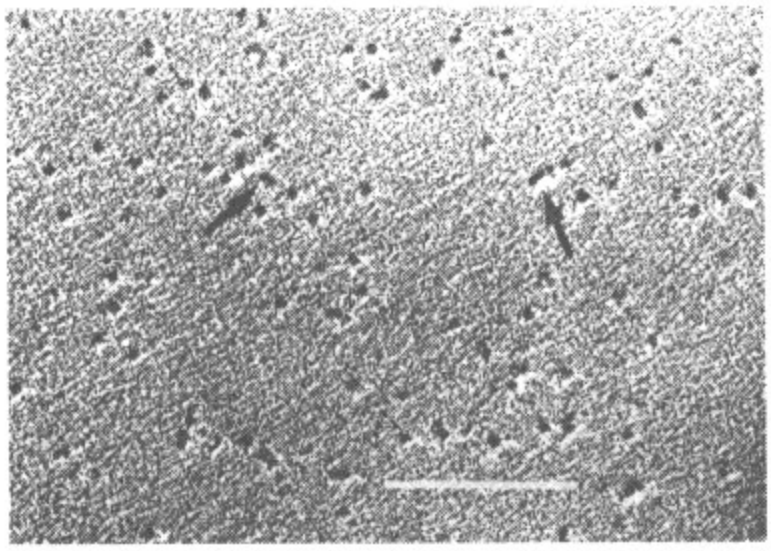
早在 1956 年，Maurice Wilkins 应用 X 射线衍射技术研究了完整细胞核内 DNA 的结构，显示有明显条带，表明有比双螺旋本身更大的重复结构。Aaron Klug、Roger Kornberg、Francis Crick 等在随后进行 X 射线衍射实验中也证实，存在间隔约为 100Å 的强重复结构，这与核小体相当，其直径约为 110Å。1974 年 Kornberg 发现，可用化学方法使组蛋白 H3 与 H4 或 H2A 与 H2B 在溶液中发生交联，此外，他还发现在溶液中 H3 和 H4 以 $(H3\text{-}H4)_2$ 四聚体形式存在。他还注意到染色质由大致等量的组蛋白和 DNA 组成，而组蛋白 H1 的浓度约为其他组蛋白的一半。这相当于每隔 200bp 的 DNA 就会有一个组蛋白八聚体（各 2 分子的 H2A、H2B、H3 和 H4）和 1 分子的组蛋白 H1。最后，他用 H3-H4 四聚体，H2A-H2B 寡聚体和 DNA 重建了染色质，并发现重建染色质的 X 射线衍射图与天然染色质相同。其他研究人员，包括 Gary Felsenfeld 和 L. A. Burgoyne 也证实，用各种核酸酶切割染色质产生大小约为 200bp 的 DNA 片段。综合上述结果，Kornberg 认为染色质是由组蛋白八聚体加 1 分子的 H1 组蛋白与约 200bp 的 DNA 结合构成的重复结构。

Georgiev 及同事发现组蛋白 H1 比另外 4 种组蛋白更易从染色质中去除。1975 年，Pierre Chambon 及同事利用组蛋白 H1 的这一特性，经胰蛋白酶和高盐缓冲液处理，选择性地从染色质中去除 H1，

发现处理后的染色质出现“串珠”结构（beads-on-a-string）（图 13.2a），他们将珠子命名为核小体。图 13.2b 显示 Chambon 及同事纯化的鸡红血细胞的核小体，并用微球菌核酸酶在珠子间切割 DNA 串。



(a)



(b)

图 13.2　核小体的早期电镜照片。（a）核小体串。Chambon 及同事用胰蛋白酶去除鸡红血细胞染色质的组蛋白 H1，显示串珠状结构。横线长度为 500nm。(b) 分离的核小体。Chambon 小组用微球菌核酸酶在核小体之间切割，然后超速离心分离这些颗粒。箭头所指为两个典型的核小体。标尺为 250nm。（*Source*：Oudet P.，M. Gross-Bellarard，and P. Chanaban，Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell* 4（1975），f. 4b & 5，pp. 286-87. Reprinted by permission of Elsevier Science.）

J. P. Baldwin 及同事对核小体进行中子散射分析（neutron-scattering analysis），该技术与 X 射线衍射技术相似，只是用中子束代替了 X 射线。样品的中子散射图为推测样品分子的三维结构提供了线索。研究人员发现了一个散射中子环，相当于约 105Å 的重复距离，与 X 射线衍射结果一致。而且，总体散射模式图显示蛋白质和 DNA 位于核小体的不同区域。基于这些结果，Baldwin 及同事认为核心组蛋白（H2A、H2B、H3、H4）形成一个球体，DNA 分子盘绕其外，这种结构有利于减少 DNA 链发生弯曲的程度。事实上，结构僵硬的双链 DNA 很难弯曲到足以紧密地结合在核小体内。他们还将组蛋白 H1 置于外部，与其易于从染色质中除去相一致。

几个研究小组都利用 X 射线衍射技术确定了组蛋白八聚体的结构。根据 Evangelos Moudrianakis 及同事 1991 年的研究结果，当从不同的方向观看时，八聚体呈现出不同的形状，但大多数视点揭示出一个由 3 部分组成的结构。如图 13.3 所示，此三分体结构包含一个 $(H3\text{-}H4)_2$ 核心附着在两个 H2A-H2B 二聚体上。其整体结构形似圆盘或饼状冰球（一面已磨损成楔型）。注意，这种结构与 Kornberg 在溶液中获得的几种组蛋白之间关系的数据是相符的，也与组蛋白八聚体可解离成一个 $(H3\text{-}H4)_2$ 四聚体及两个 H2A-H2B 二聚体的事实相一致。

DNA 到底是在哪儿装配的呢？上述实验结果不可能回答此问题，因为晶体并不包括 DNA 分子。然而，预测八聚体表面的凹槽呈现为左旋斜面，可为 DNA 提供一个缠绕的路径（图 13.3c）。1997 年，Timothy Richmond 及同事成功地获得了含有 DNA 的核小体核心颗粒晶体。正如最初所定义的，核小体含有大约 200bp 的 DNA 分子。这是经温和核酸酶处理染色质所释放出来的 DNA 片段长度。经核酸酶完全消化后产生的**核心核小体**（core nucleosome），包括 147bp DNA 片段和由 4 种**核心组蛋白**（core histone）（H2A、H2B、H3 和 H4）组成的八聚体，但是没有组蛋白 H1（通常相对容易除去）。

图 13.4 显示的是 Richmond 等确定的核小体结构。可见 DNA 分子盘绕核心组蛋白约 2 圈，H3-H4 四聚体位于顶部而 H2A-H2B 二聚体靠近底部。这种排列方式在图 13.4a 的右侧模型中特别明显。组蛋白本身的构造十分有趣，所有核心组蛋白都符合相同的**组蛋白折叠**（histone fold）基本法则。组蛋白折叠是由 2 个环所连接的 3 个 α 螺旋所组成的，而 α 螺旋都含有伸展的尾部（约占核心组蛋白量的 28%）。相对而言，尾部的结构性不强，因而在晶体结构中不包含尾部的大部分长度。去除 DNA 后，尾部就特别明显（图 13.4c）。H2B 和 H3 的尾部从 2 个相邻的 DNA 小沟所形成的间隙中穿出核心颗粒（图 13.4a 左上端的长紫色尾部）。H4 的一个尾部暴露在核心颗粒侧面（图 13.4a 右部）。此尾部富含碱性氨基酸残基，并能与相邻核小体的 H2A-H2B 二聚体的酸性区域强烈互作。这种互作可能对核小体间的交联起着重要作用，在本章后面将讨论这一部分。

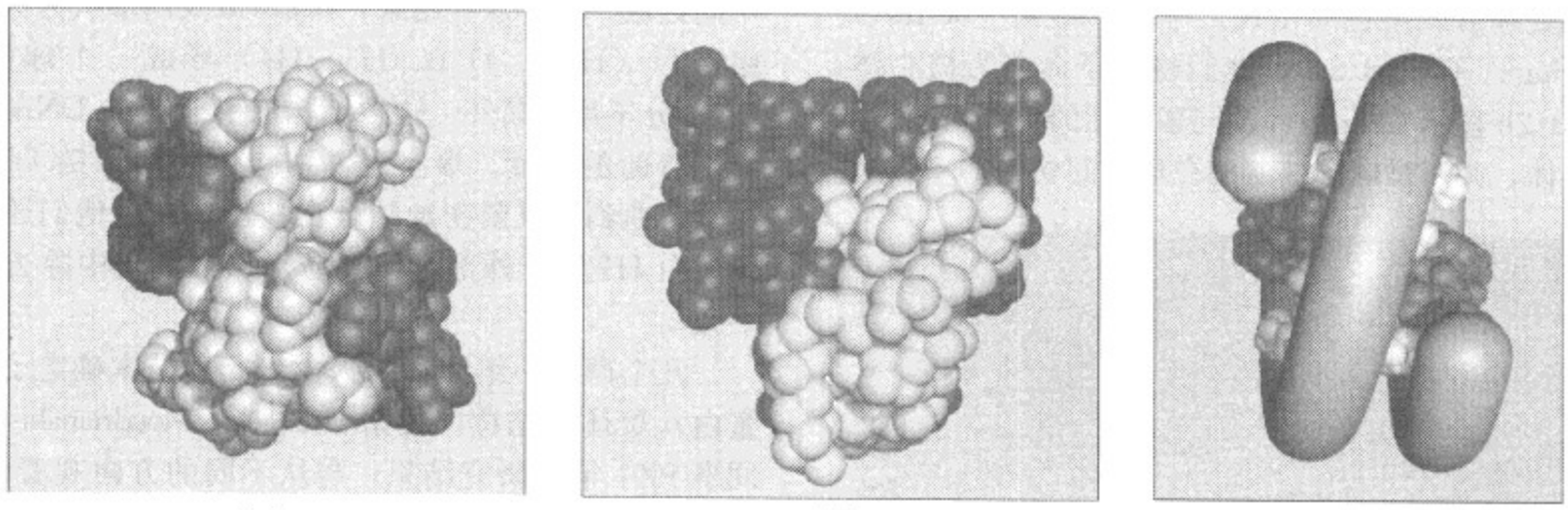


图 13.3　由 X 射线晶体学所揭示的组蛋白八聚体的两面观及推测的核小体 DNA 缠绕路径示意图。H2A-H2B 二聚体呈深蓝色，$(H3\text{-}H4)_2$ 四聚体呈浅蓝色。图（b）的八聚体是图（a）的八聚体向下旋转 90℃而成。图（a）的楔形薄边面向读者，在图（b）中则位于下方，可明显看到，楔形的缩窄部分源自 H3-H4 四聚体。(c) 预测的 DNA 盘绕组蛋白八聚体的路径。直径 20Å 的 DNA 分子（灰蓝色管状）几乎遮住了八聚体［其定位与图（a）相同］。(*Sources*：（*a-b*）Arents，A.，R. W. Burlingame，B. -C. Wang，W. B. Love，and E. N. Moudrianakis，The nucleosomal core histone octamer at 3. 1Å resolution：A tripartite protein assembly and a left-handed superhelix. *Proceedings of the National Academy of Sciences USA* 88（Nov 1991），f. 3，p. 10150.（c）Arents，A. and E. N. Moudrianakis，Topography of the histone octamer surface：Repeating structural motifs utilized in the docking of nucleosomal DNA. *Proceedings of the National Academy of Sciences UAS* 90（Nov 1993），f. 3a，1 & 4，pp. 10490-91. Copyright © National Academy of Sciences，USA.）

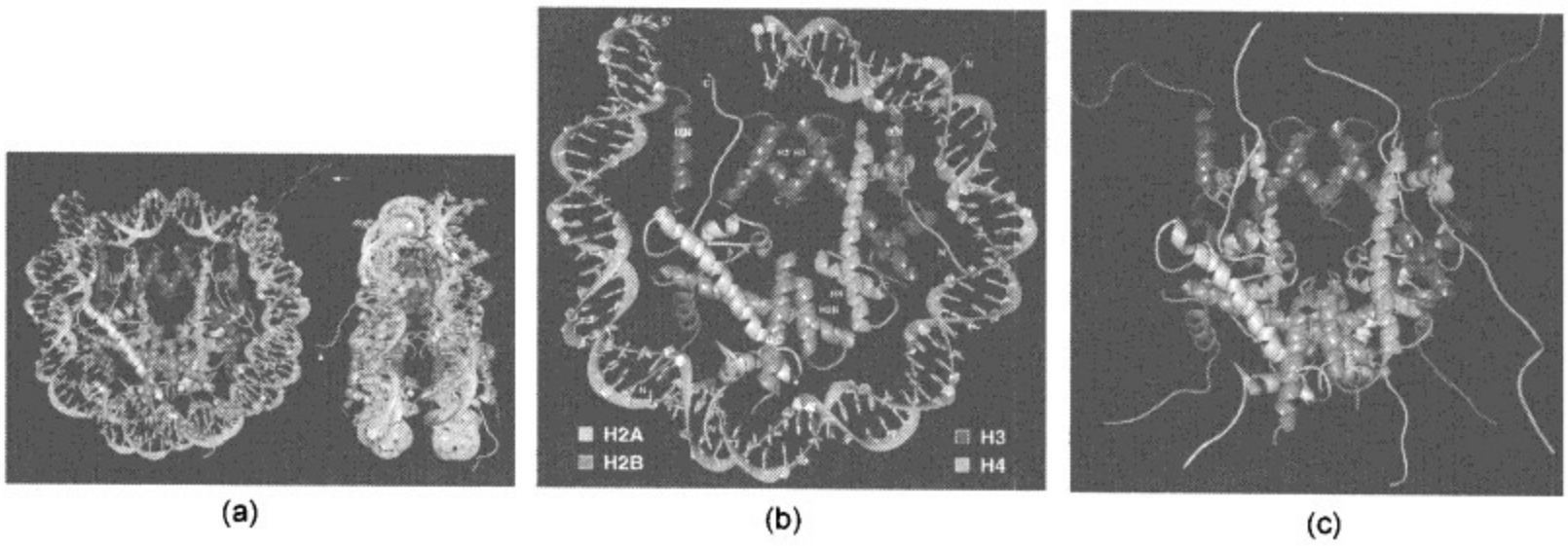


图 13.4　核小体核心颗粒的晶体结构。Richmond 等结晶了由 146bp DNA 和克隆的核心组蛋白组成的核心颗粒，而后确定了晶体结构。(a) 核心颗粒的两个视图。左边为正面，右边为侧面。外面的 DNA 分子用茶绿色表示，核心组蛋白用以下颜色：H2A，黄色；H2B，红色；H3，紫色；H4，绿色。H3 的尾部（箭头所示）通过两个相邻 DNA 分子小凹槽所形成的裂隙横穿核心颗粒。(b) 核心颗粒的一半结构，显示 73bp DNA 分子和每种核心组蛋白至少一个分子。(c) 去除 DNA 的核心颗粒。(*Sources*：（*a-b*）Luger，K.，A. W. Mäder，R. K. Richmond，D. F. Sargent，and T. J. Richmond，Crystal structure of the nucleosome core particle at 2. 8Å Resolution. *Nature* 389（18 Sep 1997）f. 1，p. 252. Copyright © Macmillan Magazines Ltd.（*c*）Rhodes，D.，Chromatin structure：The nucleosome core all wrapped up. *Nature* 389（18 Sep 1997）f. 2，p. 233. Copyright © Macmillan Magazines Ltd.）

以上核小体的模型及其他模型均揭示 DNA 分子盘绕核心颗粒约 1. 75 次，压缩 DNA 的长度 6～7 倍。1975 年，Jack Griffith 在对 SV40 微小染色体（minichromosome）的研究中也观察到 DNA 被压缩了相同的倍数。由于 SV40 DNA 在哺乳动物细胞核中复制，暴露给哺乳动物的组蛋白，形成典型的核

小体。图 13.5 显示了 SV40 DNA 的两种形态图，主图显示去除所有蛋白质后的 DNA 分子，插图显示在相同倍数下结合所有组蛋白的微小染色体。微小染色体看起来这么小的原因是在核小体中 DNA 盘绕组蛋白核心颗粒而被压缩。

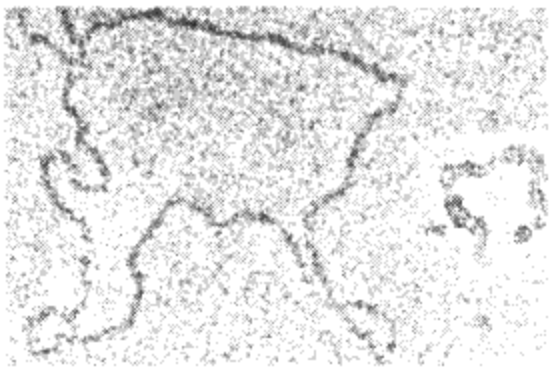


图 13.5　核小体中 DNA 的浓缩。主图是去除组蛋白的 SV40 DNA 分子，相邻的插图是 SV40 微小染色体，两者均为相同放大倍数的电镜照片。以核小体形式存在的 DNA 分子的压缩现象很明显。（*Source*：Griffith，J.，Chromatin structure：Deduced from a minichromosome. *Science* 187：1202（28 March 1975）. Copyright © AAAS.）

小结　真核生物 DNA 与叫做组蛋白的碱性蛋白质分子结合，形成核小体结构。核小体包含 4 对核心组蛋白（H2A、H2B、H3、H4），它们构成楔形盘状结构，周围包裹着约 147bp 的 DNA 片段。组蛋白 H1 比核心组蛋白更容易从核小体去除，因而不是核心核小体部分。

30nm 纤丝

在串珠状核小体结构形成之后，染色质折叠的下一级结构就是形成直径约 30nm 的纤丝（30nm fiber）。截至 2005 年，还未能结晶出染色质中比核小体核心颗粒更大的任何组分，研究人员不得不采用分辨率较低的技术，如电子显微镜（electronic microscopy，EM）来研究染色质的更高级结构。图 13.6 是电镜研究的结果，显示核小体串如何随离子强度的增加而浓缩成 30nm 纤丝。这一压缩是在核小体自身对 DNA 链浓缩（6～7 倍）的基础上又浓缩了 6～7 倍。

30nm 纤丝的结构是什么？这个问题困扰了分子生物学家数十年。1976 年，Aaron Klug 及同事在电子显微镜观察和小角度 X 射线散射数据的基础上提出了**螺线管**（solenoid）模型，即核小体排列在空心的紧密螺旋中（希腊语：solen＝管子）。但是其他人不接受螺线管模型，提出了多种其他结构模型，如核小体的 Z 字形丝带、核小体相对无序的超级珠（superbead）、核小体的不规则开放螺旋排列、双起始螺旋（two-start helix）等。其中双起始螺旋中，两个核小体间的连接体 DNA（DNA linker）在由堆叠的核小体所组成的两个螺旋排列之间 Z 字形往返排列。

为解决这个长期争论的问题，需要高分辨率的结构数据。2005 年 Richmond 及同事终于取得了突破，他们报道了四聚核小体或 4 个核小体串的 X 射线晶体结构。该结构的分辨率不高，只有 9Å，但是效果很好，足以整合高分辨率的单个核小体。图 13.7 给出了 4 个核小体的结构。图 13.7a 以一串核小体开始，此时只受 DNA 双链缠绕每个核小体的转数和核小体之间连接体 DNA 的长度所限。可以是核小体一个叠一个地将连接体 DNA 缠绕起来，或是保持 Z 字形排列形成两摞，每摞含有相隔的核小体（图 13.7b）。

实际上，四聚核小体（tetranucleosome）的晶体结构支持 Z 字形排列，其图形很复杂，图 13.7a 的示意图有助于理解，三维视图容易观察，通过以下链接：http://www.nature.com/nature/journal/v436/n7047/suppinfo/nature0368.html 可进入视频。

随着视频放映，结构不断旋转，可见所有核小体之间的连接只有 DNA，欣赏该结构的 Z 字形本质。

Z 字形结构对染色质的总体结构具有重要意义。它与以前的大多数设想包括螺线管模型都不同，但与交叉体连接的双起始螺旋一致。在该模型中两摞核小体的每一摞都形成一个左手螺旋，这种多聚核小体双螺旋的确切性质尚未由四聚核小体模型给出，但是 Richmond 及同事推测如下：首先，他们构建了一个“直接”模型，实际上就是四聚核小体一个叠加一个，但是这可引起相邻四聚核小体之间不可承受的空间干扰，于是作者又建立了一个“理想化”的模型，让摞中的每对核小体间的角度相等，这样虽然破坏了四聚核小体结构中可见的核小体间的角度，但是避免了空间干扰，产生了图 13.8 所示的合理模型。在该结构中，多聚核小体的两个螺旋很明显，在两个螺旋的某些核小体之间甚至可以看到连接体 DNA 的 Z 字形结构。

该模型留下的一个重要问题是组蛋白 H1 的位置和作用。早期的数据提示，组蛋白 H1 在形成 30nm 纤丝过程中一定起重要作用，因为没有组蛋白 H1 就不能形成纤丝。此外，化学交联实验证明组蛋白 H1 分子相对于核心组蛋白分子，更频繁地连接到其他 H1 组蛋白分子上，暗示染色质中组蛋白 H1 分子是紧靠在一起的，它们的相互作用可能在 30nm 纤丝形成过程中很重要。但其他研究发现，即使没有 30nm 纤丝，H1 分子也是紧密相连的。我们还得等待更高分辨率的结构信息来阐明组

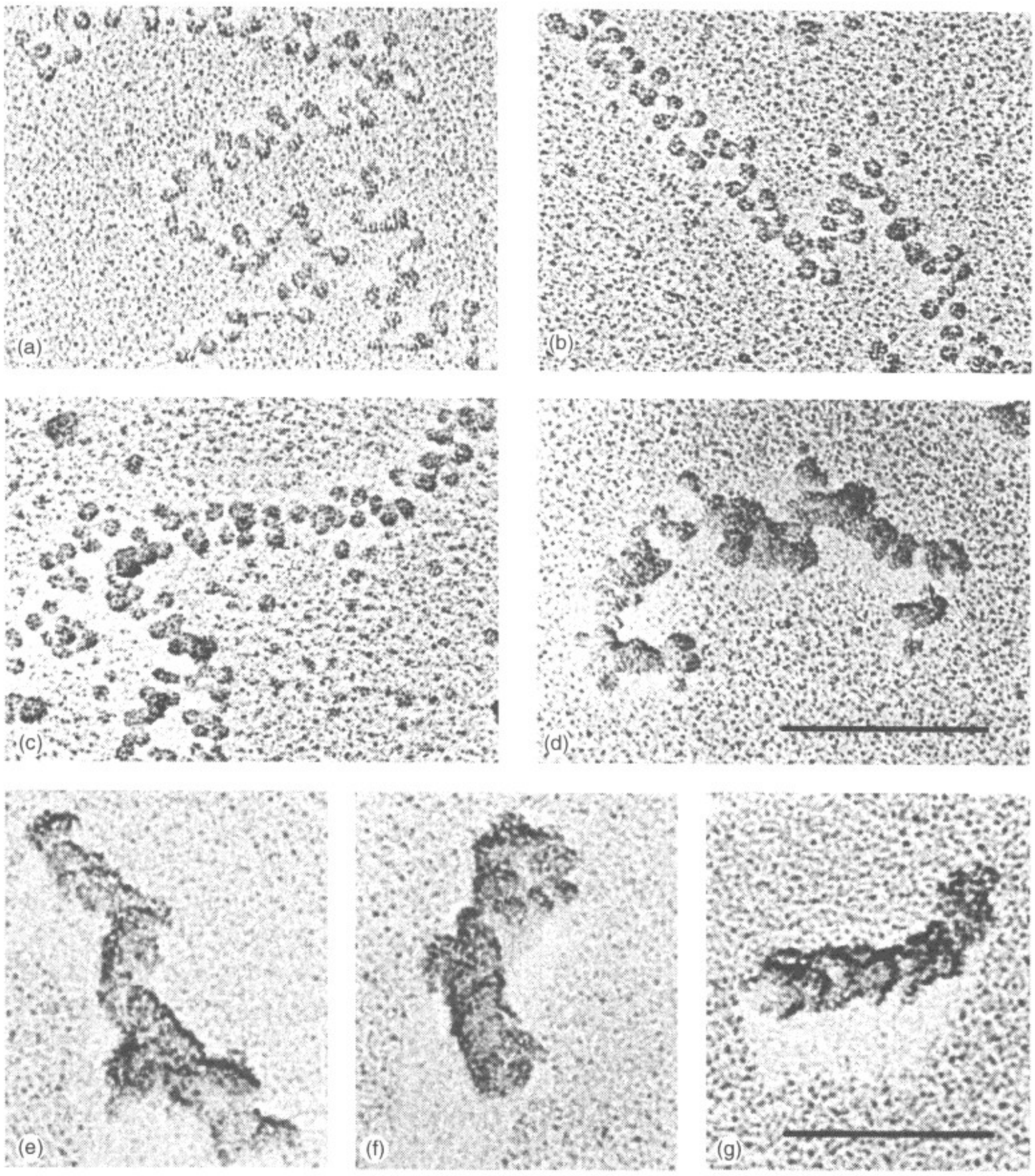


图 13.6　增加离子强度使染色质浓缩。Klug 及同事将大鼠肝脏染色质在逐步增加离子强度的缓冲液中固定进行电镜观察。图（a）～（c）为低离子强度，图（d）为中离子强度，图（e）～（g）为高离子强度。各图中的固定条件外加 0.2mmol/L EDTA 如下：（a）1mmol/L 氢氯三乙胺（TEACl）；（b）和（c）5mmol/L TEACl；（d）40mmol/L NaCl，5mmol/L TEACl；（e）～（g）100mmol/L NaCl，5mmol/L TEACl。标尺为 100nm。(*Source*: Thoma, F., T. Koller, and A. Klug, Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *Journal of Cell Biology* 83 (1979) f. 4, p. 408. Copyright © Rockefeller University Press.)

蛋白 H1 和核心组蛋白在 30nm 纤丝形成过程中的作用。

小结　四聚核小体的晶体结构由双摞核小体组成，两摞之间有 Z 字形的连接体 DNA。这种排列与螺线管模型以及多数其他关于 30nm 纤丝的模型都不同，但与交叉连接体双起始螺旋一致，即两摞核小体在左手双螺旋中可以相互缠绕。

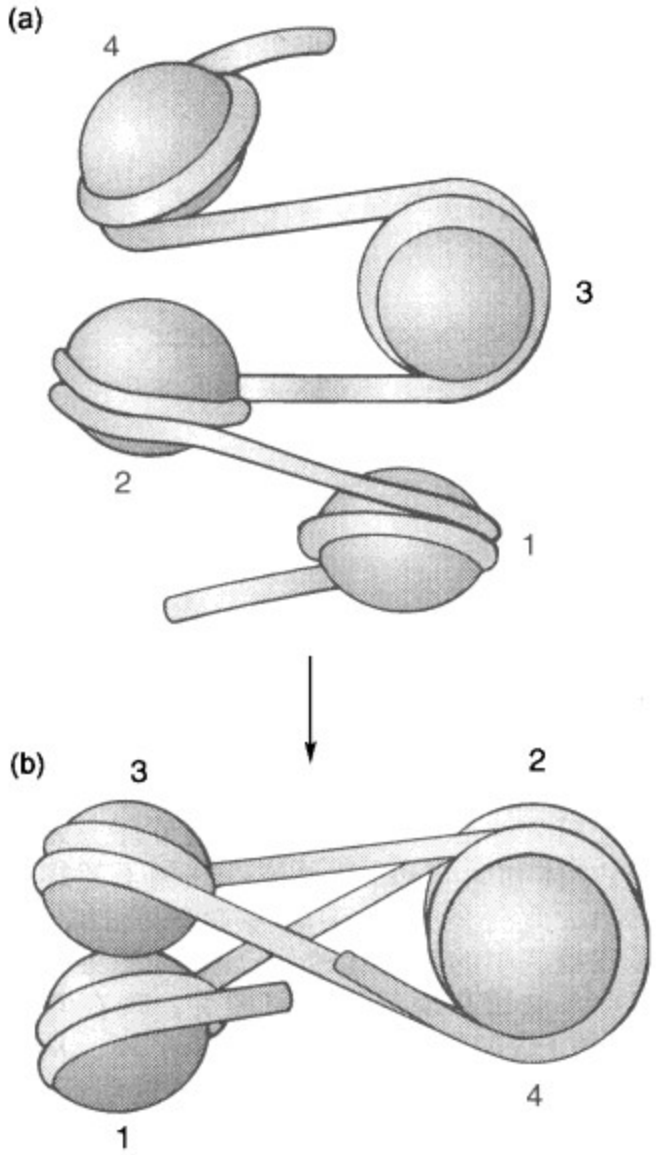


图 13.7　四聚核小体结构。两种构象的四聚核小体示意图。(a) 理论构象只受已知的 DNA 缠绕组蛋白核心的程度限制；(b) X 射线晶体学确定的构象。核小体形成两摞，连接体 DNA 在两摞的核小体之间呈 Z 字形前后折回。结果，连贯的核小体不再相邻最近，而间隔的核小体相邻最近。（*Source*: Adapted from Woodcock, G. L. *Nature Structural & Molecular Biology* 12, 2005, 1, p. 639.）

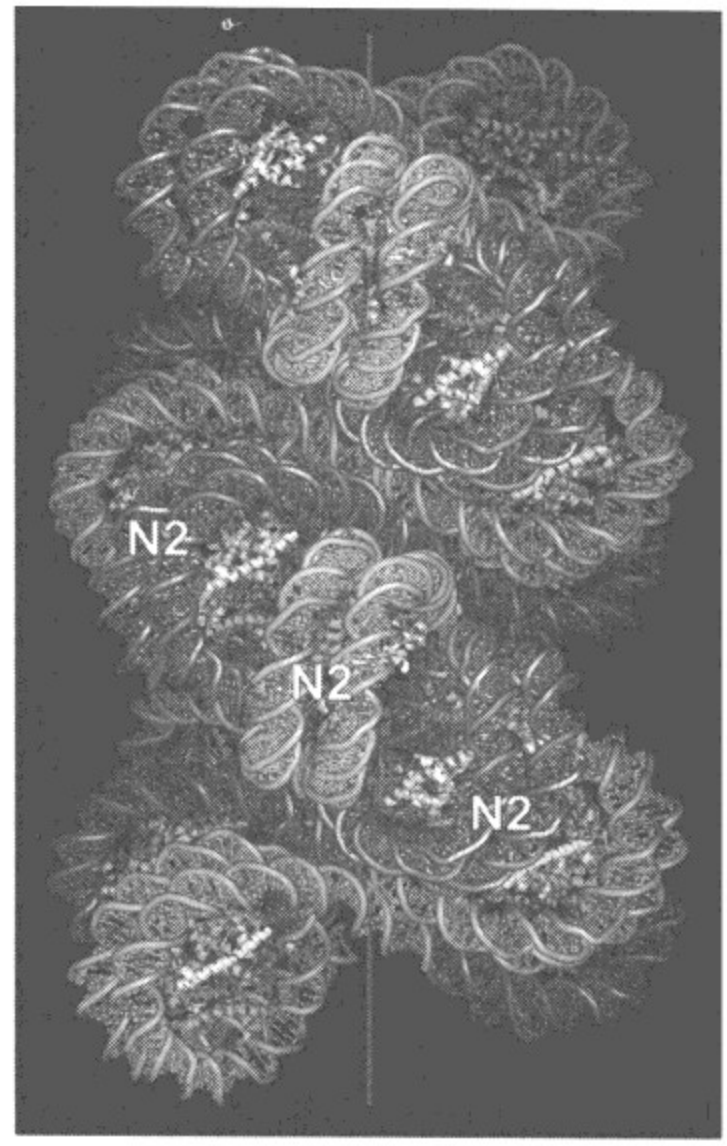


图 13.8　30nm 纤丝的模型。Richmond 及同事基于四聚核小体结构提出这个“理想化的”模型。其排列是每个核小体的二分轴（穿过核小体的中间、DNA 两个螺环之间的线）与 30nm 纤丝的轴线垂直（灰色的竖线），而且任何两个相邻核小体之间的角度都相同。(*Source*: Reprinted by permission from Macmillan Publishers Ltd: *Nature*, 436, 138-141, Thomas Schalch, Sylwia Duda, David F. Sargent and Timothy J. Richmond, “X-ray structure of a tetranucleosome and its implications for the chromatin fibre,” fig. 3, p. 140, copyright 2005.)

染色质折叠的更高级结构

在典型的间期细胞核中绝大部分染色质以 30nm 纤丝形式存在，但是进一步折叠显然是必需的，尤其是有丝分裂期的染色体，染色质高度浓缩，在光学显微镜下也清晰可见。被广泛认可的高级浓缩结构为一系列放射状的节环（loop），如图 13.9 所示。Cheeptip Benyajati 和 Abraham Worcel 在 1976 年首次提供了支持该模型的证据。他们先将果蝇染色质用 DNase I 温和消化，然后测定消化后的染色质的沉降系数。结果发现，随着消化时间的延长，沉降系数逐渐下降，直至达到稳定值。此前，Worcel 对 *E. coli* 类核（nucleoid，DNA 复合物）的研究显示出相似的结果。原因是细菌 DNA 超螺旋节环出现了越来越多的缺刻，当一个超螺旋节环被切开一个缺刻时，就会解旋成开放的大环（circle），使整个复合物的沉降系数略微降低。但是真核细胞染色体是线状的，其 DNA 如何超螺旋化呢？如果染色质纤丝像 *E. coli* 中的一样打结成环，并且在每个节环的基部紧密连接，那么每个节环就是大环的功能等价体，就能被超螺旋化。其实，DNA 在核小体上的盘绕就为其超螺旋化提供了张力。图 13.10 解释了此概念，并显示了超螺旋节环的松弛如何使该区域出现更多密度较低的染色质，从而降低了沉降系数。

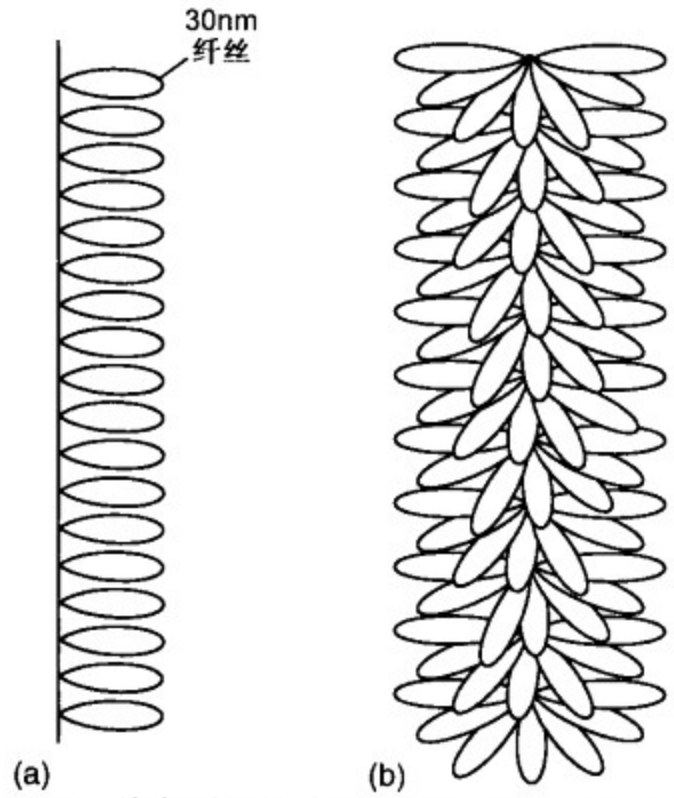


图 13.9　染色质折叠的放射节环模型。(a) 部分模型。显示部分附着在中央骨架上的染色质节环。当然，所有节环都是同一条 30nm 纤丝的一部分。(b) 较为完整的模型。显示节环如何围绕中央骨架呈三维排列。(*Source*: Adapted from Marsden, M. P. F. and U. K. Laemmli, Metaphase chromosome structure: Evidence of a radial loop model. *Cell* 17: 856, 1979.)

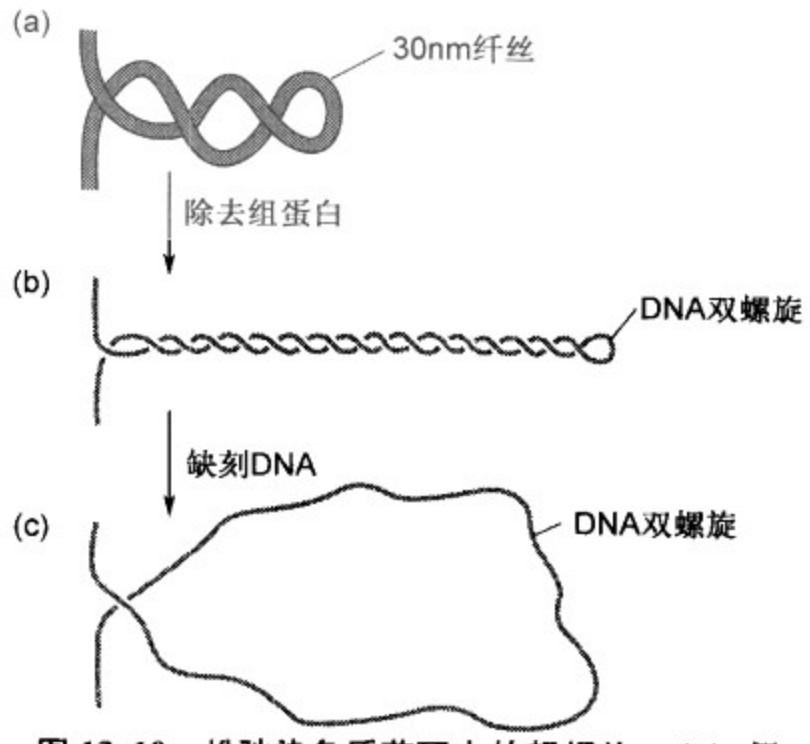


图 13.10　松弛染色质节环中的超螺旋。(a) 假设的由超螺旋节环组成的 30nm 纤丝构成的染色质节环。(b) 去除组蛋白的染色质节环。无组蛋白，核小体和 30nm 纤丝消失，只剩下超螺旋双链 DNA。此处的螺旋旋转是超螺旋的，不是 DNA 双螺旋中的普通旋转。(c) 松弛的染色质节环。DNA 被切开缺刻释放超螺旋。现在看到一个松弛的 DNA 双螺旋链形成一个节环。从 (a) 到 (c) 的每一步，染色质节环的长度明显增加，但这里未按比例画出。

节环到底有多大？Worcel 计算出果蝇染色体中每个节环约 85kb，但其他研究人员采取不同技术对脊椎动物分析的结果为 35～83kb。

图 13.11 的染色体图像也支持放射节环模型。图 13.11a 显示了人类间期染色体的边缘，其环清晰可见。图 13.11b 是人类染色体的横切图，显示 30nm 纤丝结构，放射环清晰可见。图 13.11c 展示部分去除了蛋白质的人类染色体。DNA 节环锚定在染色体骨架的中央区域。所有这些图像有力地支持了染色体中放射节环状纤丝的概念。

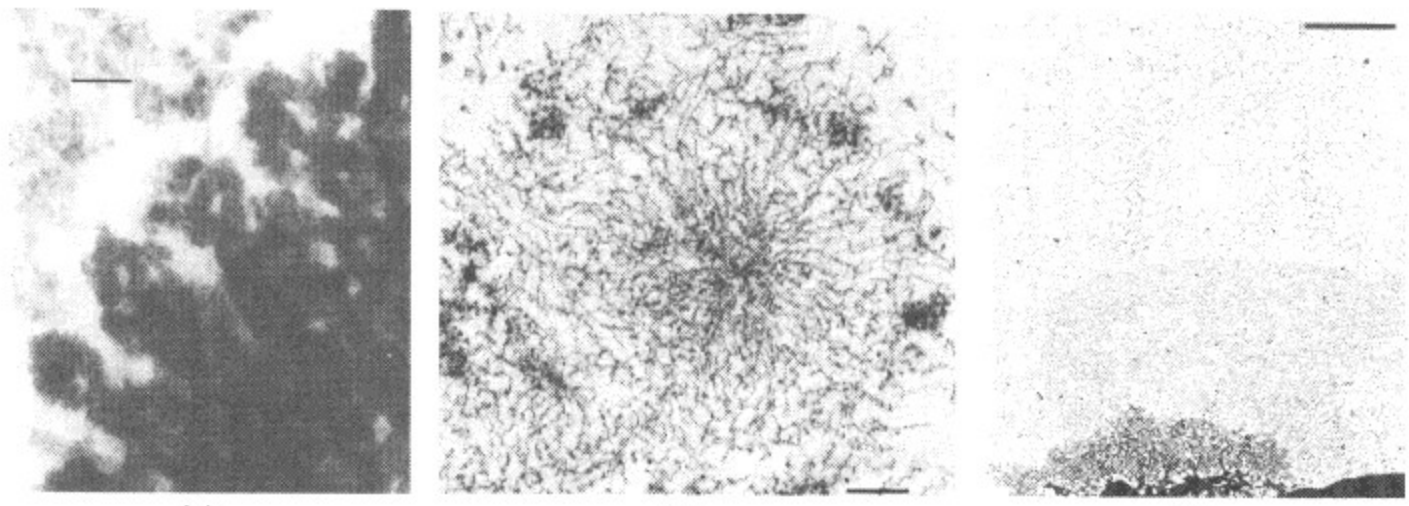


图 13.11　人类染色体节环的 3 种不同视图。(a) 用己烯-乙二醇分离的人类染色体边缘扫描投射电镜图。标尺为 100nm。(b) EDTA 处理后发散的人类染色体横切面的透射电镜图。观察到的染色质纤维是 30nm 核小体纤丝。标尺为 200nm。(c) 去蛋白质的人类染色体透射电镜图。显示从中央骨架发散的 DNA 节环。标尺为 2000nm。(*Sources*: (a) Marsden, M. P. F. and U. K. Laemmli, Metaphase chromosome structure: Evidence for a radial loop model. *Cell* 177 (Aug 1979) f. 5, p. 855. Reprinted by permission of Elsevier Science. (b) Marsden and Laemmli, *Cell* 17 (Aug 1979) f. 1, p. 851. Reprinted by permission of Elsevier Science. (c) Paulson, J. R. and U. K. Laemmli, The structure of histone depleted metaphase chromosomes. *Cell* 12 (1977) f. 5, p. 823. Reprinted by permission of Elsevier Science.)

小结 采用染色体沉降和透射电镜技术，证实了真核染色体的放射节环状结构。30nm 纤丝可能通过某种折叠或螺旋化形成 50nm 纤丝，该纤丝随后形成 35～85kb 的环，附着在染色体的中央骨架上。

13.3 染色质结构与基因活性

对组蛋白作为基因活性的重要调控因子的关注几经反复。当首先确认体外添加组蛋白可关闭转录时，分子生物学家异常兴奋。而当阐明组蛋白在染色质结构中的作用时，多数研究者转而关注组蛋白在结构上的作用，而忽视其作为基因活性调控因子的作用，只把组蛋白看成 DNA 分子的骨架。现在，我们转了一大圈又反过来重新研究组蛋白的调控功能了。

组蛋白对 5S rRNA 基因转录的影响

早在 1984 年，Donald Brown 及同事就证明，组蛋白特别是组蛋白 H1，在体外具有抑制特定基因活性的效应。他们主要研究了非洲爪蟾 5S rRNA 基因，该基因在 RNA 聚合酶 III 及转录因子 TFIIIA、TFIIIB、TFIIIC 的协助下进行转录。

在非洲爪蟾单倍体细胞中 5S rRNA 基因数量巨大，约为 20 000 个，组成两个不同的家族。第一个家族占全部基因的 98%左右，因为它们仅在卵母细胞中转录，故称为**卵母细胞 5S rRNA 基因**（oocyte 5S rRNA gene）。较小的家族只有约 400 个基因，可在卵母细胞和体细胞中转录，故称为**体细胞 5S rRNA 基因**（somatic 5S rRNA gene）。这些差异产生了一个关键问题：是什么引起卵母细胞基因在卵母细胞中有活性，而在体细胞中无活性的？

核小体的一个功能可假设如下：在体细胞中卵母细胞基因或至少其启动子与核小体形成复合物，因而处于抑制状态，而在卵母细胞中，核小体相对较少，从而基因具有活性。事实上，Brown 小组找到了证据支持这一假说。

Brown 及同事首先转录了分别来自卵母细胞和体细胞的纯化的 5S 基因，发现在 3 种转录因子的协助下，RNA 聚合酶 III 能很好地转录这两种基因。接着，他们从卵母细胞和体细胞中温和地提取染色质，然后尝试在体外对二者进行转录，所得结果有很大差异：在体外卵母细胞 5S 基因在卵母细胞染色质中有活性，但在体细胞染色质中没有活性，表明在体细胞染色质中存在某些物质抑制卵母细胞基因的表达。

正如我们刚才设想的，“某些物质”与核小体和组蛋白相关。Brown 及同事还发现，在某些体细胞染色质制备物中，卵母细胞 5S 基因意外地保持有活性。为查明此意外活性的原因，他们查看了有活性和无活性染色质中组蛋白的组成。无活性染色质含有全部 5 种组蛋白，而活性染色质中缺失组蛋白 H1（图 13.12a）。盐处理或离子交换层析可在体细胞染色质中重新激活卵母细胞 5S rRNA 基因，而这些处理同样选择性地去除了组蛋白 H1（图 13.12b）。

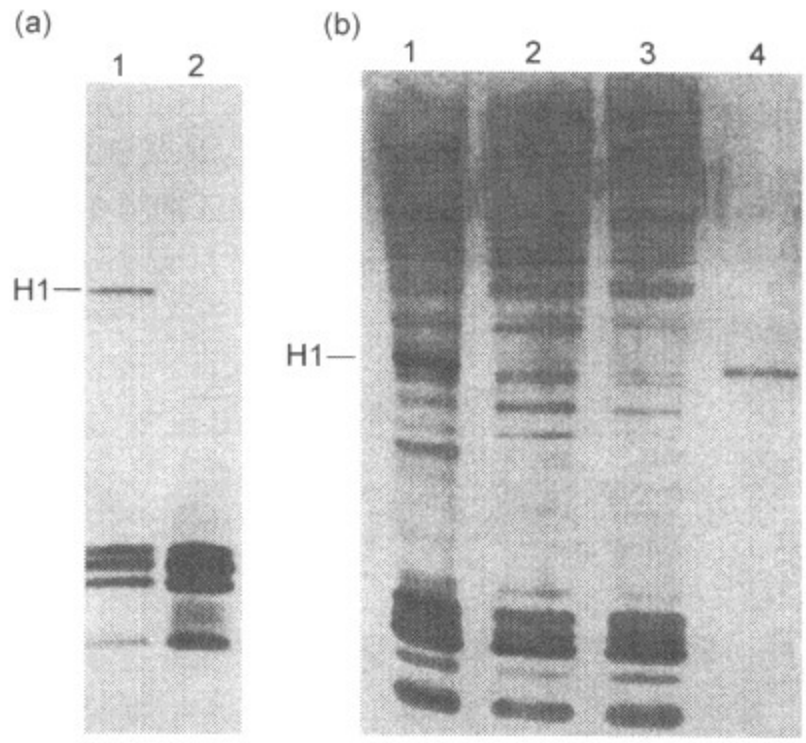


图 13.12 有活性和无活性染色质的组蛋白组成。(a)“偶然”解阻遏染色质的组蛋白。Brown 及同事用凝胶电泳检测了体细胞染色质中组蛋白的含量。泳道 1 含有被抑制染色质中的组蛋白，泳道 2 含有纯化过程中“偶然”解阻遏的染色质的组蛋白，可能是组蛋白的部分蛋白水解。后一种染色质缺乏组蛋白 H1。(b) 处理染色质的组蛋白。Brown 及同事对体细胞染色质做不同处理后，进行电泳：泳道 1，未处理；泳道 2，0.6mol/L KCl 冲洗；泳道 3，离子交换层析；泳道 4，含有一种组蛋白 H1 标签。在两个处理的样品中均缺乏组蛋白 H1。（*Source*：Schliessel, M. S. and D. D. Brown, The transcriptional regulation of *Xenopus* 5S RNA genes in chromatin: The roles of active stable transcription complexes and histone H1. *Cell* 37 (July 1984) f. 3b, 4. pp. 905-906. Reprinted by permission of Elsevier Science.）

如果将组蛋白 H1 再加回到体细胞染色质中结果会怎么样？Brown 及同事这样做了，并得到以下结果：当组蛋白 H1 的浓度达到每 200bp DNA 有一分子的组蛋白 H1 时，即染色质中的天然水平，5S rRNA 的合成显著降低，剩余的转录都是体细胞 5S rRNA 基因所贡献的。

对上述结果令人信服的解释是，体细胞含有转录因子（TFIIIA、TFIIIB 和 TFIIIC），它们能够与体细胞 5S rRNA 基因形成稳定的前起始复合物，但不易与卵母细胞 5S 基因形成复合物。因此，核小体就在卵母细胞基因（包含启动子在内）上形成，而组蛋白 H1 将这些核小体交联在一个有序排列的阵列中（可能在 30nm 纤丝中），并使其保持抑制状态。相比之下，作用于体细胞的转录因子防止核小体形成，或至少防止其形成有序的交联结构。因此，这些基因保持有活性。换句话说，转录因子与组蛋白在同一位点存在竞争关系，正如图 13.13 所描述的。如果转录因子占上风，则基因保持活性状态，若组蛋白占优势，它们将基因的调控区捆绑在交联的核小体中并抑制基因的表达。

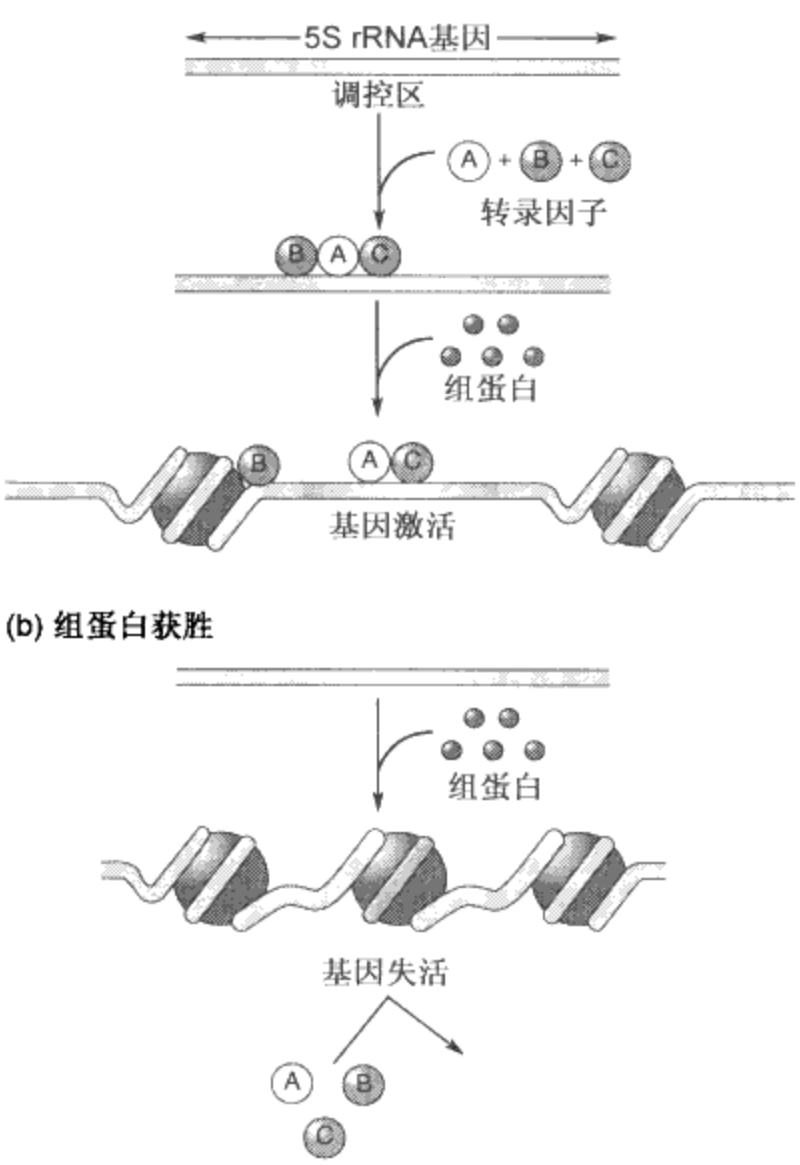


图 13.13　转录因子和组蛋白 H1 对 5S rRNA 调控区的竞争。（a）转录因子 TFIIIA、TFIIIB、TFIIIC 获胜，并结合到调控区（紫色）。组蛋白也许仍可以在调控区以外形成，但不形成交联。（b）组蛋白获胜，与调控区结合，将调控区捆绑在交联的核小体上。基因无法接触转录因子，保持失活状态。

如果这个模型是正确的，我们就可用盐从体细胞染色质中除去组蛋白 H1，然后加入来自卵母细胞的转录因子，最后重新加入组蛋白 H1。加入转录因子是为了保证其首先与基因结合并使基因激活。Brown 等的实验确实证明卵母细胞基因被激活了。

小结　非洲爪蟾中 5S rRNA 基因的两个家族分别是卵母细胞基因和体细胞基因。前者仅在卵母细胞中表达，后者在卵母细胞和体细胞中均可表达。造成这种差异的主要原因在于体细胞基因与转录因子形成更为稳定的复合物。转录因子通过阻止核小体与体细胞基因内部调控区形成稳定复合物而使基因保持活性。形成稳定的抑制复合物需要组蛋白 H1 的参与，一旦此类复合物形成，转录因子就不再结合，基因活性受到抑制。

组蛋白对 II 类基因转录的影响

James Kadonaga 及同事证明组蛋白与 III 类基因相互作用的原则也适用于 II 类基因与组蛋白的相互作用。

核心组蛋白　1991 年，Paul Laybourne 和 Kadonaga 对体外核心组蛋白和组蛋白 H1 对 RNA 聚合酶 II 转录的影响进行了细致的对比研究。结果发现核心组蛋白（H2A、H2B、H3、H4）与克隆的 DNA 形成核心核小体，并对基因活性产生温和抑制效应（约为 4 倍）。转录因子对该抑制作用没有影响。当他们在核心组蛋白基础上再添加组蛋白 H1 时，抑制作用极大增强（为 25～100 倍），而且激活因子能阻止这种抑制作用。在此方面，这些因子与另一些因子（TFIIIA、TFIIIB、TFIIIC）相似，后者能与组蛋白 H1 竞争非洲爪蟾的 5S rRNA 基因的调控区。

Laybourne 和 Kadonaga 设计的实验是在有或无激活因子（已知会影响克隆基因的转录）条件下以含有目的基因的质粒 DNA 和组蛋白重建染色质。同时，他们还加入拓扑异构酶 I 以保持 DNA 的解旋状态。然后利用引物延伸检测重建的染色质能否被核提取物所转录。第一个实验只加核心组蛋白而没有组蛋白 H1。他们以质量比为 0.8～1.0 的比例添加一定量的组蛋白到 DNA 中，这个浓度足以在每 200bp 的 DNA 处形成一个核小体。

他们利用含有果蝇 *Krüppel* 基因的重建染色质，证明果蝇的核提取物能转录 *Krüppel* 基因（图 13.14）。但是当核心组蛋白产生核小体的量达到每 200bp DNA 一个核小体的密度时（达到其生理水平），导致转录的部分抑制（泳道 2 与泳道 5 相比，相当于对照值的 25%）。值得注意的是，用这种方法检测到该基因的转录起始位点具有较高的异质性，所以我们看到一系列引物延伸产物。

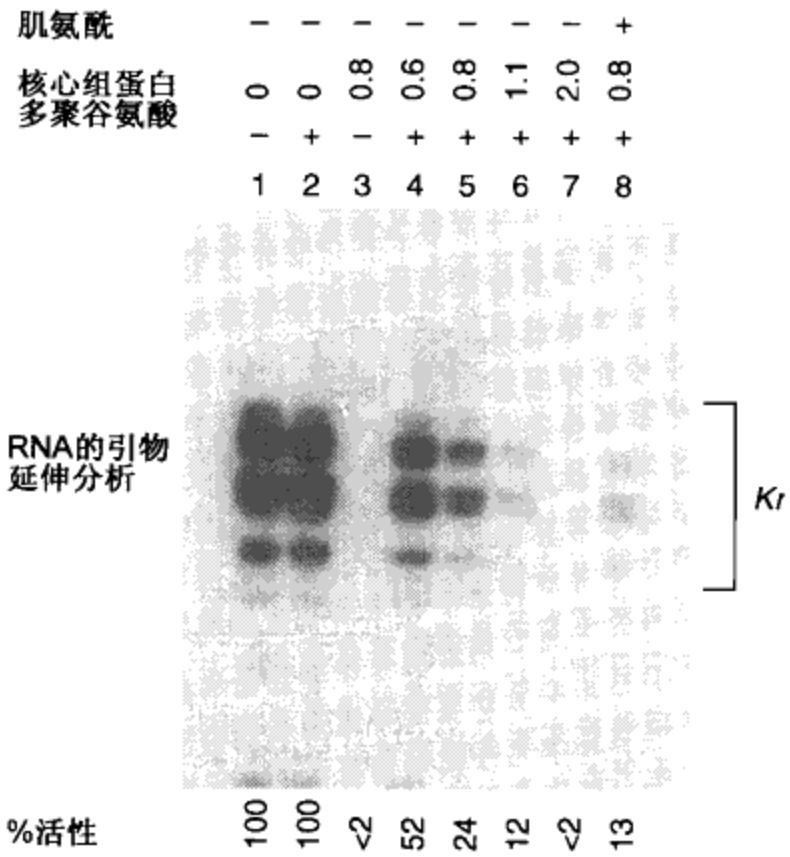


图 13.14　重建染色质的体外转录。Laybourne 和 Kadonaga 利用含有果蝇 *Krüppel* 基因的质粒 DNA 和核心组蛋白按蛋白质与 DNA 不同比例重构了染色质，如图顶部所示。然后用引物延伸实验检测转录效率。对应于 *Krüppel* 基因转录的不同信号在右侧括号标出。泳道 1，裸 DNA；泳道 2，裸 DNA 加多聚谷氨酸（作为组蛋白返回 DNA 的载体）；泳道 3～7，不同比例组蛋白-DNA 的染色质；泳道 8，肌氨酰用于阻止再起始，只能发生一轮转录。核心组蛋白对 *Krüppel* 基因转录的抑制是剂量依赖性的。（*Source*：Laybourn，P. J. and J. T. Kadonaga，Role of nucleosomal cores and histone H1 in regulation of transcription by RNA polymerase II. *Science* 254（11 Oct 1991）f. 2B，p. 239. Copyright © AAAS.）

对实验中核心组蛋白抑制转录活性达到 75%的结果，作者给出了两种可能的解释。其一，核小体只能减缓全部 RNA 聚合酶催化进程的 75%，不能完全抑制。其二，75%的聚合酶完全被核小体抑制，而仍有 25%的启动子游离于核小体之外，因此 RNA 聚合酶能够与之结合起始转录。对照实验表明剩余的 25%转录可以通过限制酶切割紧接转录起始位点的下游区域而被消除。这一事实说明，该位点未被核小体所结合，因而证实了假说的正确性。

小结　核心组蛋白（H2A、H2B、H3、H4）能在裸 DNA 上组装成核小体核心颗粒。当重建染色质达到每 200bp DNA 一个核小体核心颗粒的密度时，75%的转录将被抑制（相对于裸 DNA）。剩余 25%的转录活性由启动子区未被核小体核心颗粒覆盖所引起。

组蛋白 H1　基于组蛋白 H1 作为核小体交联子（cross-linker）的推测，我们期望组蛋白 H1 会增强重建染色质中由核心组蛋白引起的转录抑制作用。正如 Laybourne 和 Kadonaga 所证明的，结果确实如此。他们用含有两种增强子-启动子结构的 DNA 重建染色质：①pG_5E4（5 个 GAL4-结合位点偶联腺病毒 E4 最小启动子）；②pSV-Kr（SV40 早期启动子的 6 个 GC 框偶联果蝇 *Krüppel* 基因的最小启动子）。实验中，不仅添加了核心组蛋白，还有不同数量的组蛋白 H1，比率为每个核心核小体 0～1.5 个组蛋白 H1 分子，然后体外转录重建的染色质。

图 13.15 中的奇数泳道显示随组蛋白 H1 量的增加，引起模板转录活性的逐渐丧失，直到检测不到转录。但是在中等水平的组蛋白 H1 中（每核心组蛋白有 0.5 分子），激活因子能阻止大部分的抑制作用。例如，在 pG_5E4 质粒重建染色质中，与 GAL4 结合位点作用的杂合激活因子 GAL4-VP16 使模板活性提高了 200 倍。部分增强作用（8 倍）是激活因子的刺激作用所致，甚至在裸 DNA 中也观察得到。其余 25 倍显然是**抗阻遏作用**（antirepression）产生的，即对组蛋白抑制作用的阻止。同样，以 pSV-Kr 启动子 DNA 重建染色质，激活因子 Sp1（与启动子中的 GC 框结合），能使模板转录活性提高 92 倍。若以裸 DNA 为模板，Sp1 仅能使其活性提高 2.8 倍，因此其他的 33 倍则是由于抗阻遏作用产生。在第 12 章我们已经讨论过确切的激活成分是什么，在那里的实验是以裸 DNA 为模板进行转录分析的。

这些结果与图 13.16 的模型一致。该模型中，组蛋白 H1 通过与两个核小体之间的连接体 DNA（正好含有一个转录起始位点）结合而引起抑制。如果激活因子（绿椭圆形）与组蛋白 H1 同时加入，它就能阻止转录抑制效应。但是，这些因子不能逆转已形成的核小体核心颗粒的作用（甚至在组蛋白 H1 不存在的情况下）。当激活因子（紫色长方形）遇到核小体阻抑启动子的情况，它就与染色质重建因子（chromatin-remodeling factor）合作，至少在核小体未与组蛋白 H1 交联时，将核小体推至一边。

Kadonaga 与同事也研究了另一种蛋白 GAGA 因子（GAGA factor），它与 *Krüppel* 启动子和其他果蝇启动子的富含 GA 的序列结合。该因子自身没有激活转录活性，实际上还稍微抑制转录，但在组蛋白之前添加 DNA 时，GAGA 因子可阻止组蛋白 H1 的抑制作用，从而引起转录速率的显著净增加。因此，GAGA 因子好像是个纯的抗阻遏物，而不像我们已经学过的典型激活因子同时具有抗阻抑和转录激活活性。

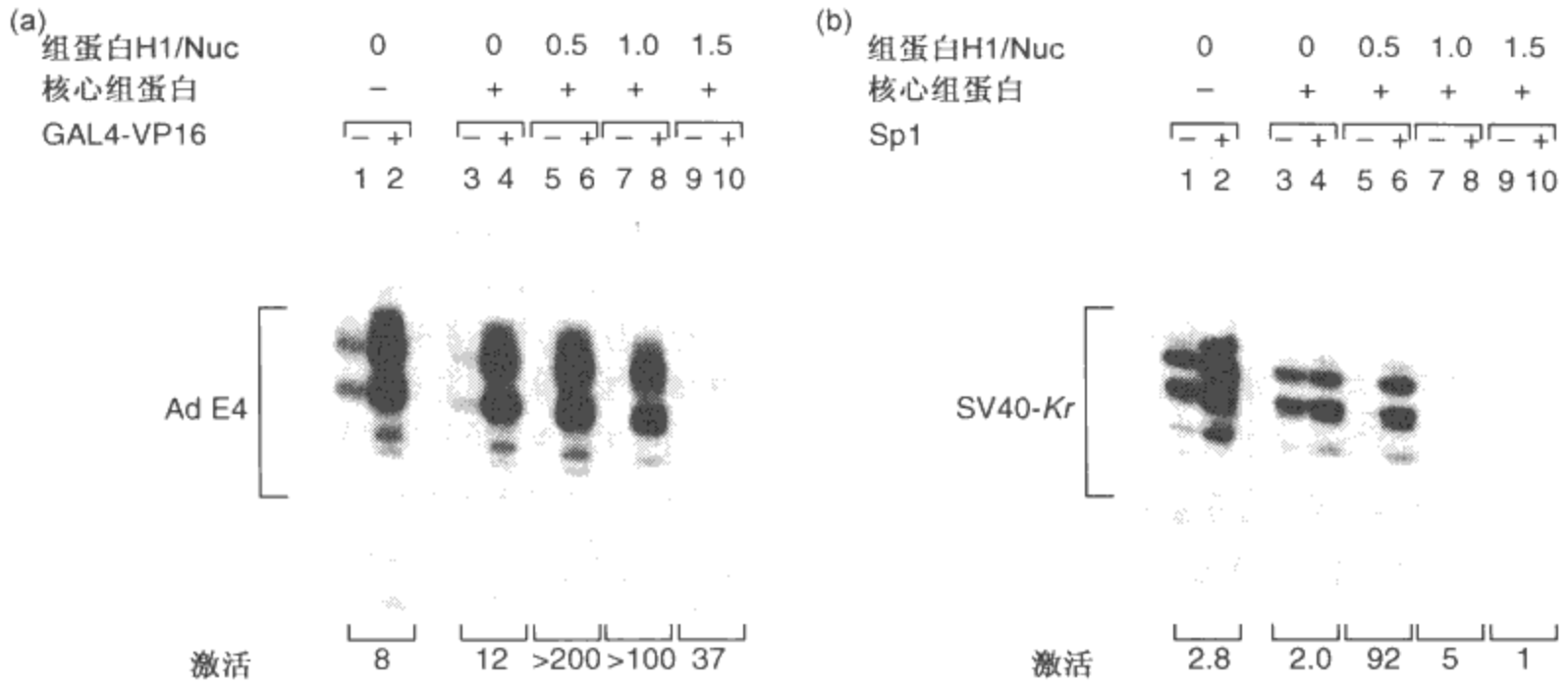


图 13.15　组蛋白与激活因子对转录的竞争效应。Laybourne 和 Kadonaga 在有或无核心组蛋白和组蛋白 H1 条件下重建了染色质，如图上端所示。然后用引物延伸实验检测激活因子存在与否对转录的影响。每个激活因子激活转录的程度在泳道下方给出。每个激活因子的真实激活作用见两个图的泳道 1 和 2，以裸 DNA 为模板。其他泳道均以染色质为模板，其中明显的激活作用是由抗阻抑效果引起。(a) GAL4-VP16 的效果。染色质含有腺病毒 E4 启动子，其上有 5 个 GAL4 结合位点。与 E4 转录相关的信号在左侧括号标出。(b) Sp1 的效果。染色质含有 *Krüppel* 启动子外加 SV40 启动子的 GC 盒，GC 盒对 Sp1 响应。与 *Krüppel* 转录相关的信号在左侧标出。(*Source*: Laybourn, P. J. and J. T. Kadonaga, Role of nucleosomal cores and histone H1 in regulation of transcription by RNA polymerase II. *Science* 254 (11 Oct 1991) f. 7, p. 243. Copyright © AAAS.)

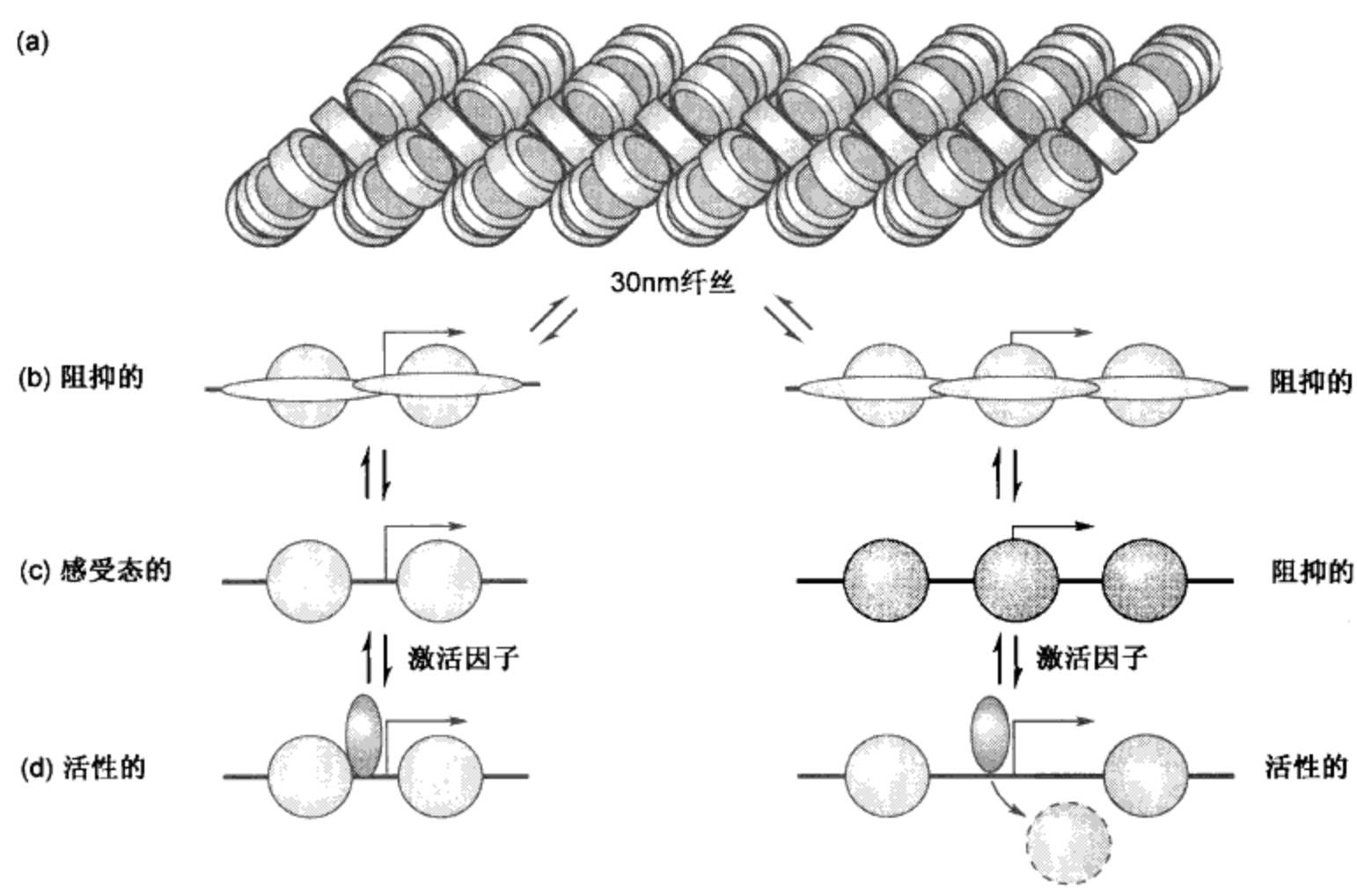


图 13.16　转录激活模型。(a) 从顶部的 30nm 纤丝开始。(b) 30nm 纤丝可打开，有两种被抑制的染色质：右侧的与核小体（蓝色）交联覆盖着启动子，导致阻抑。左侧的无核小体覆盖启动子，但组蛋白 H1（黄色）交联核小体在启动子两翼与核小体交联，仍然抑制基因。(c) 除去组蛋白 H1，获得两种状态的染色质：左侧的，启动子暴露，基因准备转录；右侧的，核小体仍覆盖启动子，基因保持抑制状态。(d) 抗阻抑。如果基因调控区未被核小体阻断（左侧），激活因子（绿色）可与之结合，并与其他因子一起启动转录。如果基因调控区被一个或多个核小体结合（右侧），激活因子（紫色）协同其他因子，包括染色质重建因子将核小体移开，启动转录。(*Source*: Adapted from Laybourn, P. J. and J. T. Kadonaga, Role of nucleosomal cores and histone H1 in regulation of transcription by polymerase II. *Science* 254: 243, 1991.)

小结 组蛋白 H1 在核心核小体抑制的基础上进一步抑制模板活性，这种抑制功能可被转录因子所抵消。有些因子，如 Sp1 和 GAL4，既可作为抗阻遏物（阻止组蛋白 H1 引起的抑制），也能作为转录激活因子，有的只是抗阻遏物，如 GAGA 因子。

核小体定位

图 13.16 的激活和抗阻抑模型说明，转录因子通过去除与启动子结合的核小体或阻止其率先与启动子结合，而引起抗阻抑效应。这两种模式都蕴涵了**核小体定位**（nucleosome positioning）的概念，即激活因子迫使核小体只能占据启动子的外围位置，而不能进入其中。

无核小体区 一系列证据表明在活跃基因的调控区存在**无核小体区**（nucleosome-free zone）。M. Yaniv 及同事对 SV40 病毒 DNA 调控区做了一个图解实验。如本章前述，在感染的哺乳动物细胞中，SV40 DNA 以微小染色体形式存在。Yaniv 注意到某些活跃转录的 SV40 微小染色体在感染晚期出现一个明显的无核小体区（图 13.17）。我们会预测这种无核小体区至少包含一个晚期启动子。事实上，SV40 的早期和晚期启动子相距很近，二者只间隔 72bp 重复的增强子。这就是无核小体区吗？问题是环状染色体没有起点和终点，如果没有某种标记，就无法判断我们所看到的是环的哪一部分。Yaniv 及同事利用限制酶酶切位点作为标记。已知 *Bgl*I 位点靠近调控区的一末端，*Bam*HI 和 *Eco*RI 位点在环的其他位点上，如图 13.18a 所示。因此，若无核小体区包含调控区域，*Bgl*I 就能在该区内切割，而其他两个限制酶只能在较远位点切割（图 13.18b）。图 13.19 的实验显示 *Bam*HI 和 *Bgl*I 酶切产生的正是预期结果，*Eco*RI 酶切也符合预期结果（未显示）。

我们甚至可以看到 *Bgl*I 在无核小体区非对称切割，因为它在线性化的微小染色体一端留下了长长的无核小体尾，另一端则没有。如果无核小体区对应于 SV40 的一个启动子，该启动子相对于 *Bgl*I 位点不对称排列，那么这一结果就是我们所期望的。如果无核小体区对应病毒复制起点；该复制起点跟 *Bgl*I 位点几乎重合，则以上结果就不是我们所期望的。

DNase 超敏感性 DNA 无核小体区的另一个特征是对 DNase 的超敏感性。活跃转录的染色质区属**DNase 敏感区**（DNase-sensitive）（比主体染色质敏感约 10 倍）。而活跃基因调控区为 **DNase 超敏感区**（DNase-hypersensitive）（比主体染色质敏感约 100 倍）。与我们所预测的一样，SV40 DNA 调控区属 DNase 超敏感区。Yaniv 证实了这一点，他从感染 SV40 病毒的猴细胞中提取染色质，用 DNase I 温和

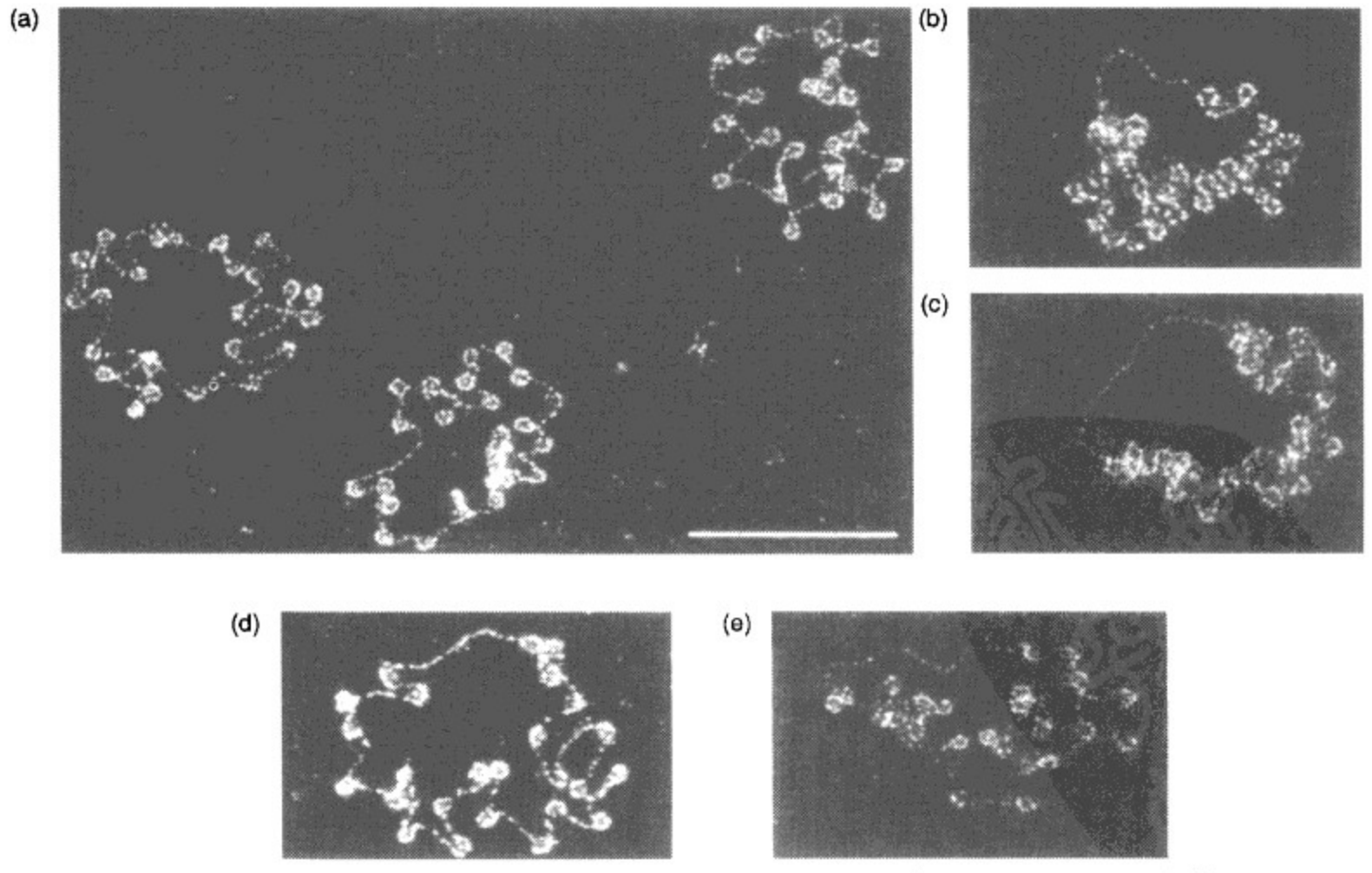


图 13.17 SV40 微小染色体的无核小体区。（a）没有大范围无核小体区的 3 个微小染色体；（b）～（e）有明显的无核小体区的 4 个 SV40 微小染色体。标尺为 100nm。（*Source*：Saragosti，S.，G，Moyne，and M，Yaniv，Absence of nucleosomes in a fraction of SV40 chromatin between the origin of replication and the region coding for the late leader RNA. *Cell* 20（May 1980）f. 2，p. 67. Reprinted by permission of Elsevier Science.）

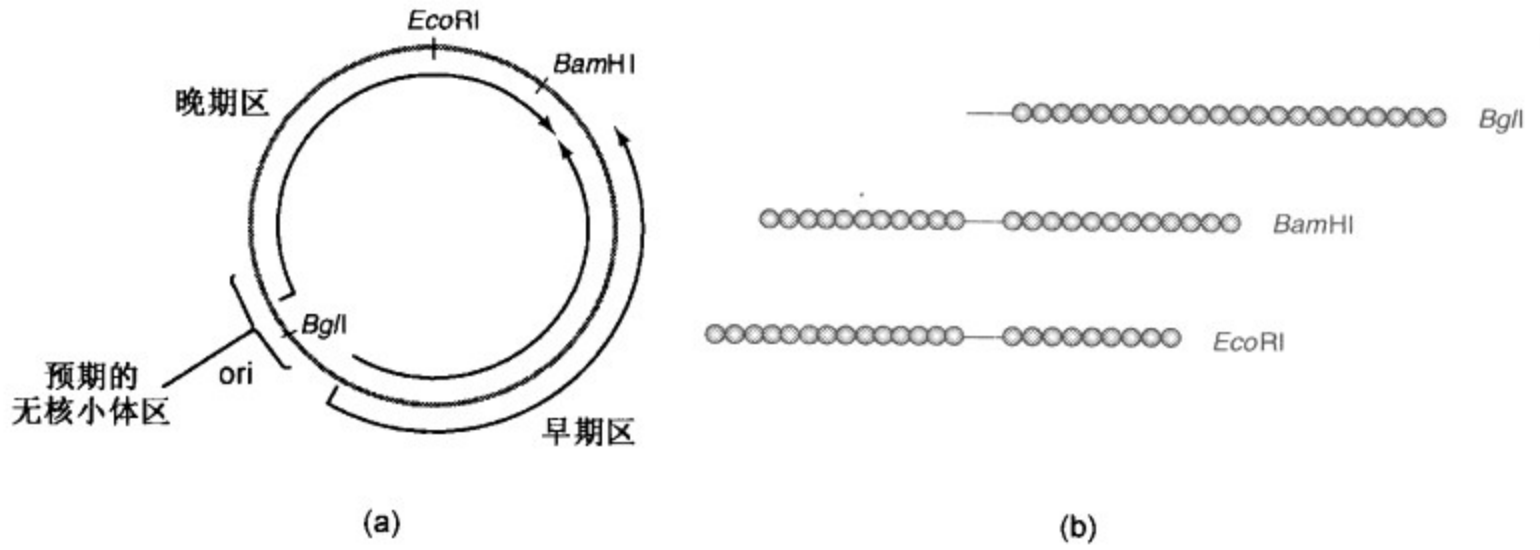


图 13.18　定位 SV40 微小染色体的无核小体区的实验程序。（a）SV40 基因组图谱显示 3 种限制酶酶切位点 *Bgl*I、*Bam*HI、*Eco*RI。调控区包围着复制起点（ori），晚期调控区为顺时针方向；（b）三种限制酶切割感染晚期细胞的微小染色体的预测结果。假设晚期调控区为无核小体，3 个酶均切一次使微小染色体线性化。预测 *Bgl*I 在无核小体区的一端切割，因而产生一端有无核小体区的微小染色体。预测 *Bam*HI 在无核小体区的对面位点酶切，产生无核小体区位于中间的微小染色体。同样，*Eco*RI 产生无核小体区位于非对称位置的微小染色体。

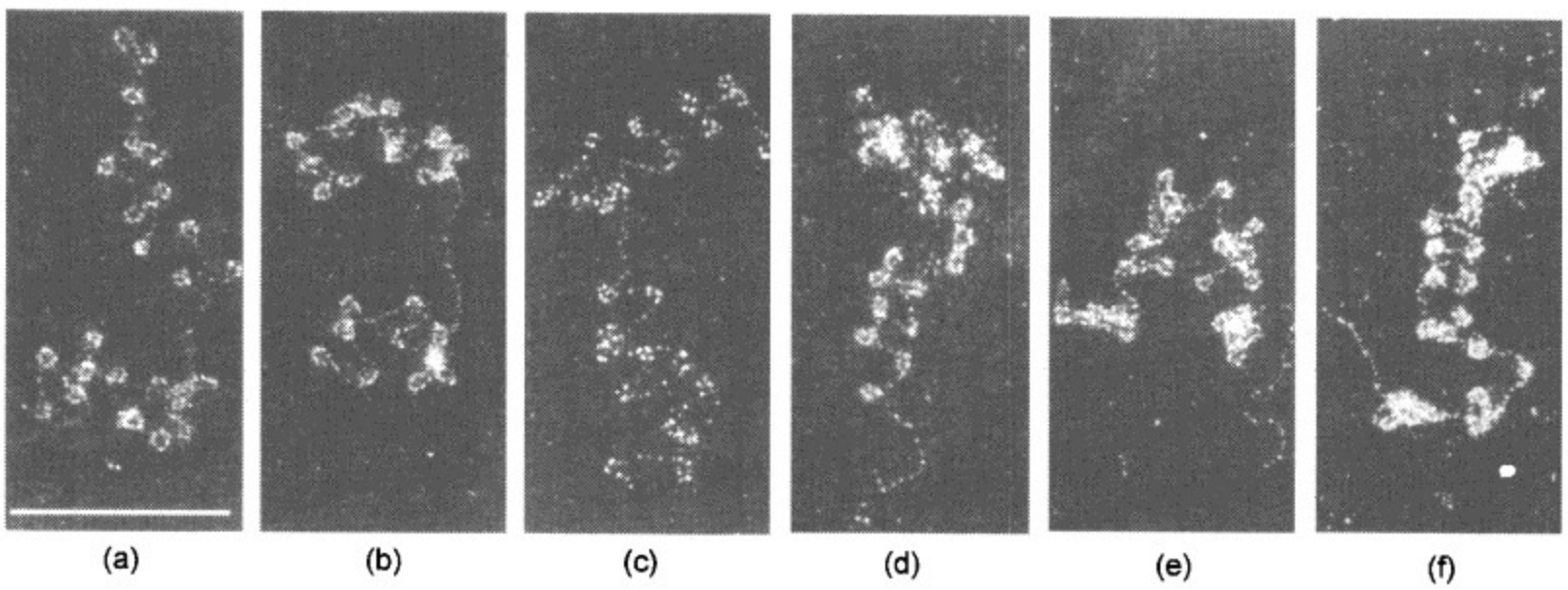


图 13.19　对 SV40 微小染色体上无核小体区的定位。Yaniv 及同事用 *Bam*HI（图 a～c）或 *Bgl*I（图 d～f）酶切感染后期细胞的 SV40 微小核小体。正如图 13.18 所预期的，*Bam*HI 产生的无核小体区位于微小染色体中央，而 *Bgl*I 产生的无核小体区位于微小染色体的末端。标尺为 100nm。

消化后，纯化 SV40 DNA，再经 *Eco*RI 酶切、电泳，最后用放射性 SV40 DNA 作探针进行 Southern 印迹检测。

图 13.18a 显示 *Eco*RI 和 *Bgl*I 位点在环状染色体上相距 67%和 33%。因此，如果靠近 *Bgl*I 位点的无核小体区真是 DNase 超敏感区，那么，DNase 就能在此处切割，而 *Eco*RI 将在其对应的特异位点酶切，产生包括 SV40 整个基因组的 67%和 33%的 2 个片段。实际上，如图 13.20 所证明的，在病毒感染细胞 24h、34h、44h 后所进行的酶切实验，都产生了大量 67%片段，少量 33%片段及一些更短的片段。这说明 DNase I 确实在围绕 *Bgl*I 位点的较小区域内切割染色质，无核小体区和 DNase 超敏感区相吻合。

活跃基因调控区的 DNase 超敏性是一种普遍现象。例如，红细胞中 ε-球蛋白基因的 5′-侧翼区是 DNase 超敏区。事实上，球蛋白基因的 DNase 超敏性是对任意给定时间内基因活性的很好指示。

图 13.21 阐明了应用 Southern 印迹检测 DNase 超敏感性基因的原理。图上方的（a）和（b）显示核小体在活跃和非活跃基因上的排列、限制性核酸内切酶（RE）两个识别位点的位置。如果用DNaseI 轻微消化含非活跃基因的染色质，则什么也不会发生，因为没有 DNase 超敏感位点。而同样方法处理

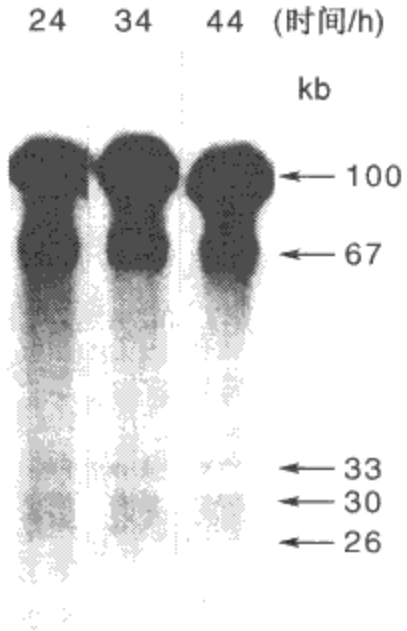


图 13.20　定位 SV40 微小染色体中的 DNase 超敏感区。Yaniv 及同事分别在 SV40 病毒感染猴细胞 24h、34h、44h 后提取细胞核，用 DNase I 处理，再经 *Eco*RI 酶切、电泳、Southern 印迹分析 DNA 产物。因为 *Eco*RI 在环状基因组中无核小体区顺时针方向的 33%的位置酶切，预期出现两种产物：33%和 67% SV40 基因组总长度的片段，这里假定 DNase 超敏感区和无核小体区重合。实际上，67%的片段量很大，而部分 33%片段已被降解成更小的片段。因此，DNase 超敏感区确实与无核小体区相对应，而此无核小体区足够大以至于产生了一系列降解的 DNA 产物。(*Source*: Saragosti, S., G. Moyne, and M. Yaniv, Absence of nucleosomes in a fraction of SV40 chromatin between the origin of replication and the region coding for the late leader RNA. *Cell* 20 (May 1980) f. 7, p. 71. Reprinted by permission of Elsevier Science.)

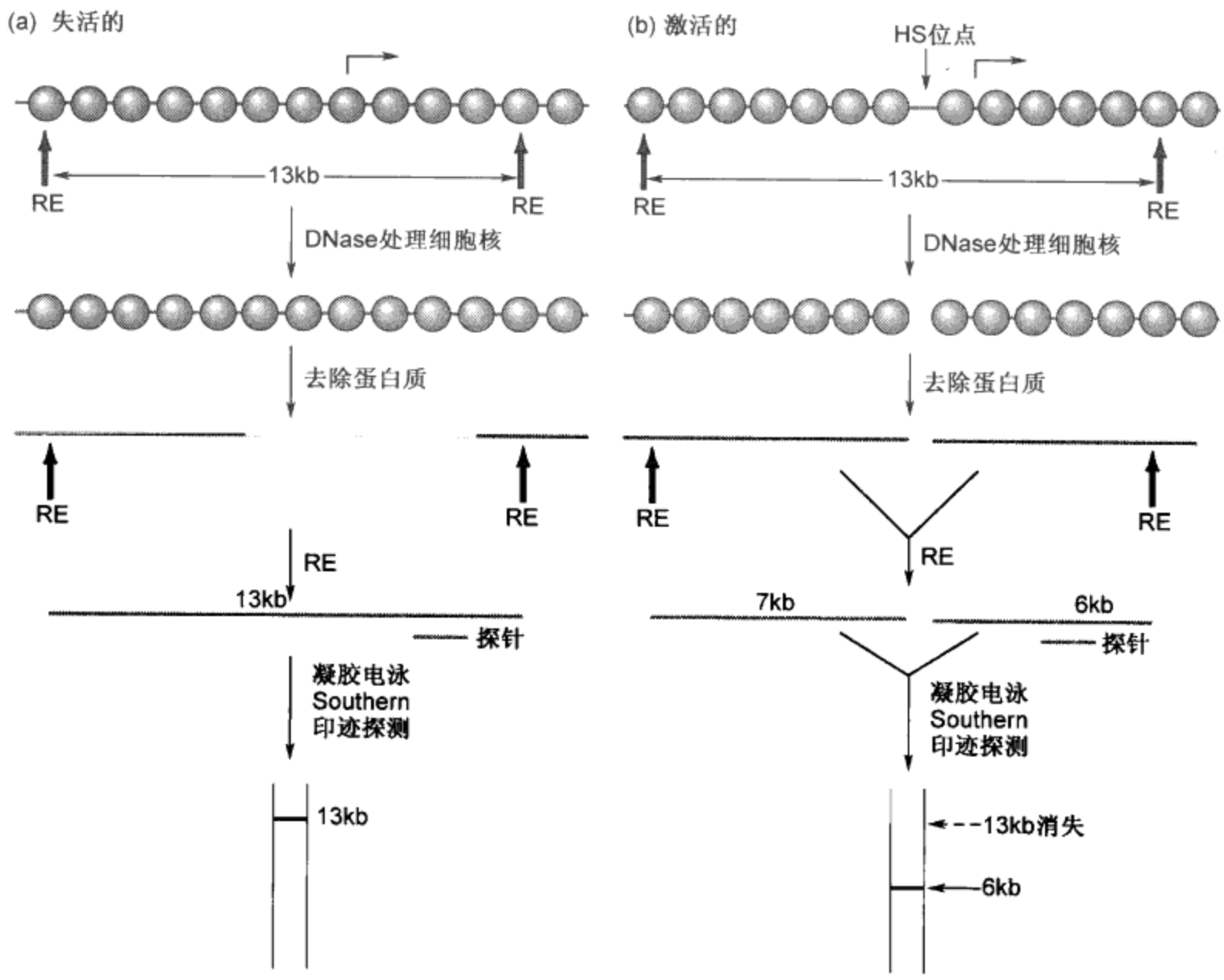


图 13.21　检测 DNase 超敏感区的实验策略。(a) 非活跃基因，无 DNase 超敏感性。基因及其调控区域与核小体复合，用 DNase I 温和处理含该基因的细胞核，DNA 不降解。接着，从这些细胞核提取 DNA，除去蛋白质，用限制性核酸内切酶 (RE) 消化，产生包括基因调控区的 13kb 的 DNA 片段。电泳 RE 酶切产物，对片段做 Southern 印迹膜，用基因特异性探针检测。结果出现 13kb 片段。(b) 活跃基因，有 DNase 超敏感性。活跃基因有一至多个无核小体区，可能与启动子、增强子、隔离子或另一个调控区对应。因此，当用 DNaseI 温和处理含有活跃基因的细胞核时，其超敏感位点 (HS 位点) 被消化。接着，分离 DNA，去除蛋白质，用限制性核酸内切酶消化，电泳片段，对片段做 Southern 印迹膜，同图 (a) 进行探针检测。由于 DNase 酶切，13kb 片段消失，但在 6kb 处出现一个新片段。7kb 片段不能与探针杂交所以检测不到。这个实验揭示在下游 RE 位点的上游约 6kb 处有一个 DNase 超敏感位点。实际操作中，常常增加 DNase 的浓度，使 13kb 条带亮度逐渐下降，而 6kb 条带亮度逐渐增加。

具有活跃基因的染色质，DNase 将攻击启动子附近的超敏感位点。现在，将蛋白质从两种 DNA 分子中除去，然后用 RE 酶切。限制性片段经电泳、转 Southern 印迹膜、用短的基因特异性探针检测斑点（绿色）。来自非活跃染色质的 DNA 是完整的，所以经 RE 酶切后产生了一个能与探针杂交的 13kb 片段。但是来自活跃染色质的 DNA 由于包含 DNase 超敏感位点，经 DNase 和 RE 消化，产生了两个片段（6kb、7kb）。其中 6kb 片段可被探针检测到，但 7kb 片段不能。更长时间的 DNaseI 消化处理后，13kb 片段通常会消失。

图 13.22 显示 1987 年 Frank Grosveld 及同事所做的关于人球蛋白基因簇的实验结果。该基因簇包括 5 个活跃状态的球蛋白基因，顺序为 $5'$-ε-^{G}γ-^{A}γ-δ-β-$3'$。Grosveld 及同事注意到，当 β-球蛋白基因单独转入转基因小鼠时（第 5 章），其功能最多只有其正常水平的 10%。当 β-球蛋白基因插入某些位点时，其功能明显强于其他位点。他们推测是 β-球蛋白基因之外的某些物质在调控其表达效率。事实上，好几个位点都影响这一效率，而且这些位点都是 DNase 超敏感性的。

其中 5 个位点（1、2、3a、3b、4）位于 ε-球蛋白基因位点的上游，如图 13.22e 所示。Grosveld 及同事按图 13.21 所述分析了 DNase 超敏感位点。图 13.22e 显示了 3 种不同探针（*Eco*RI、*Eco*Bgl、

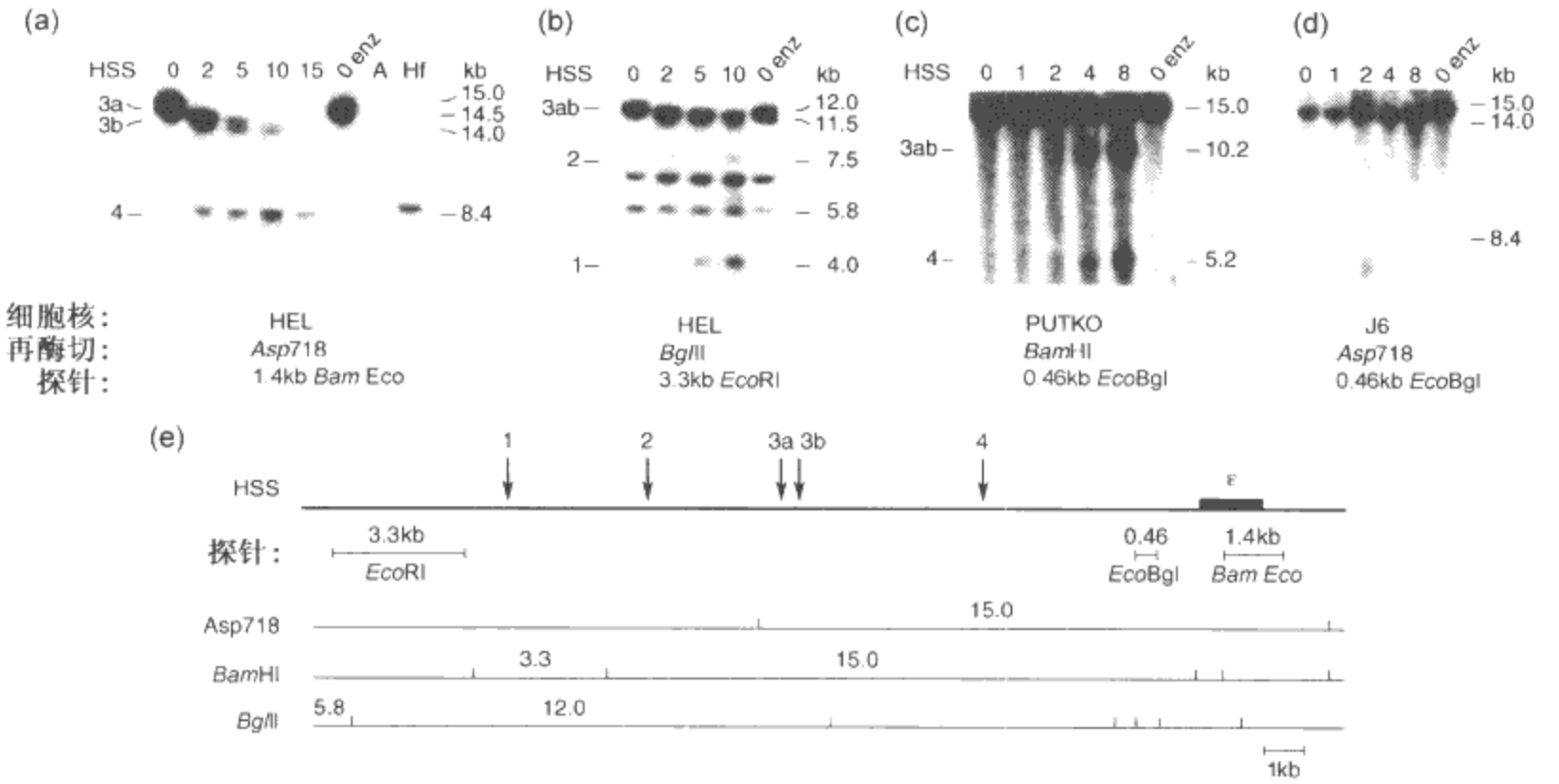


图 13.22　人类球蛋白基因 $5'$-侧翼区 DNase 超敏感位点的作图。图（a）～（d）是 Grosveld 及同事处理 HEL、PUTKO、J6 细胞核的结果，如图下“Nuclei”所示，以低浓度 DNaseI 处理不同时间（分钟）无酶切对照（0 enz）如图上部所示。然后从细胞核提取 DNA，用蛋白酶 K 去除蛋白质，用不同的限制酶（底部“Recut”所示）切割 DNA，电泳 DNA 片段，转膜，用探针检测（底部“Probe”所示）。与切割超敏感位点（HSS）1、2、3a、3b、4 对应的 DNA 片段见图左侧。图（a）中，标有 A 和 Hf 的泳道是用 *Alu*I 或 *Hin*fI 替代 DNaseI 切割 DNA 的结果。(e) 人类 ε-球蛋白基因位点 $5'$-侧翼区图谱，显示 3 种探针的位置和图（a）～（d）中所用 3 种限制性核酸内切酶的酶切位点。(*Source*: Reprinted from *Cell* v. 51, Grosveld et al., p. 976. © 1987, with permission from Elsevier Science.)

BamEco）的位置。Grosveld 及同事对细胞核进行处理，一组是表达 β-球蛋白基因的两个人类细胞系——白血病（HEL）细胞和另一种人类红细胞系（PUTKO），另一组是不能表达 β-球蛋白基因的细胞系——人类 T 细胞（J6）。图中的“0 enz”泳道代表无 DNaseI 处理的结果，其他数字标示的泳道是 DNaseI 不同处理时间的结果。

图 13.22a 显示 HEL 细胞的结果，限制酶是 *Asp*718，探针是 1.4kb *Bam* Eco，可清楚看到 DNaseI 切割在位点 3a、3b、4。为了检测基因上游更远处的超敏感位点，Grosveld 及同事使用了 3.3kb *Eco*RI 探针，如图 13.22b 所示，这一次，可观察到切割位点位于位点 1、2、3a、3b 处，但位点 2 的切割滞后且相对较弱。与 5.8kb DNA 片段对应的 5.8kb 条带与探针发生反应，如图 13.22e。6.8kb 条带是与不相关基因非特异性杂交的结果，可通过提高杂交的严谨性而消除。图 13.22e 是 PUTKO 细胞的结果，限制酶是 *Bam*HI，探针是 0.46 *Eco*Bgl。可以看到位点 3a、

3b、4 的切割。Grosveld 及同事用同样的方法检测了 β-球蛋白基因下游的另一个 DNase 超敏感位点。

最后，他们对 J6 T 细胞的 DNase 超敏感性进行了检测。J6 T 细胞没有活跃的球蛋白基因，如图 13.22d 所示，未检测到 DNase 超敏感性。这一结果支持这个假说：超敏感性与基因专一性因子的存在相对应，该因子能将核小体排斥于活跃基因而不是非活跃基因之外。

Grosveld 等推测这些位点与最佳外源基因表达所需的重要的基因调控区相对应。确实，当把整个球蛋白基因簇，包括这些位点，移植到转基因小鼠时，β-球蛋白基因与内在的 β-球蛋白基因一样有效地表达，不论该基因插入小鼠基因组的什么位置都具有活性。这些实验定义了一个重要的调控区，我们现在称为球蛋白**基因位点调控区**（locus control region，LCR）。

小结　活跃基因具有 DNase 超敏感调控区。这种超敏感性至少部分是由于缺乏核小体产生的。

组蛋白乙酰化

1964 年 Vincent Allfrey 发现了组蛋白的乙酰化和非乙酰化两种形式。乙酰化发生在赖氨酸侧链的氨基基团。Allfrey 还证明组蛋白的乙酰化与基因的活性有关，即非乙酰化组蛋白抑制 DNA 转录，而乙酰化组蛋白却是弱转录抑制因子。这些结果说明细胞核中的某些酶可使组蛋白发生乙酰化和去乙酰化，从而影响基因的活性。为了证实这个假说，需要鉴定出这些酶。然而，30 多年来这些酶仍然难以确定，部分原因是它们在细胞中的量很低。

终于，在 1996 年 James Brownell 和 David Allis 成功地纯化和鉴定了一种**组蛋白乙酰转移酶**（histone acetyltransferase，HAT）。该酶从供体（乙酰辅酶 A，acetyl-CoA）上将乙酰基团转移到核心组蛋白上。他们采用一个创造性的策略分离了该酶：以四膜虫（*Tetrohymena*，纤毛原生动物）细胞为材料，因为四膜虫含有高度乙酰化的组蛋白，说明其细胞中含有相对高浓度的 HAT。他们从大核（四膜虫的大细胞核含有活跃转录的基因）中制备提取物，然后在灌有组蛋白的 SDS 胶中进行电泳。为了检测 HAT 的活性，他们将凝胶浸泡在乙酰 CoA 溶液中（乙酰基团被放射性标记）。如果胶体中包含有 HAT 活性的条带，HAT 就能将标记的乙酰基团从乙酰 CoA 转移到组蛋白上，从而在凝胶上有 HAT 活性的位置出现一条被标记的乙酰化组蛋白条带。为检测标记的组蛋白，将未结合的乙酰 CoA 冲洗掉，然后对凝胶进行荧光照相。结果如图 13.23 所示，HAT 活性条带与 55kDa 蛋白两者在位置上相对应。Brownell 和 Allis 把该蛋白质命名为 p55。

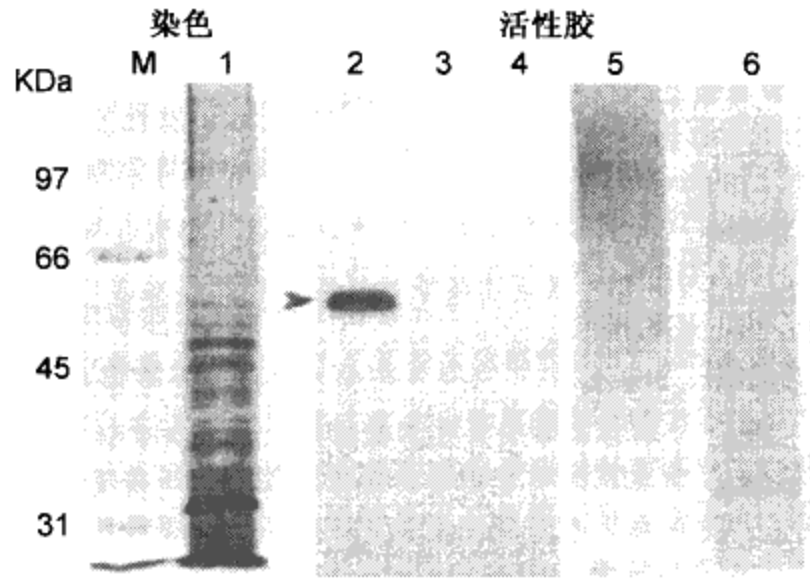


图 13.23　活性胶分析组蛋白乙酰转移酶（HAT）活性。Brownell 和 Allis 将四膜虫大核提取物分别在含有组蛋白（泳道 2～4）、小牛血清白蛋白（BSA，泳道 5）或无蛋白质（泳道 1、6）条件下进行 SDS-PAGE。然后银染检测蛋白质（泳道 M、1），或用 ^{3}H 标记乙酰基的乙酰 CoA 处理以检测 HAT 活性。洗去未结合的乙酰 CoA，荧光照相检测 ^{3}H 标记的乙酰基团。泳道 2 显示清晰的 ^{3}H-乙酰化组蛋白条带，表明有 HAT 活性存在。泳道 3 和 4 未显示活性，因为电泳前细胞核提取物被加热（泳道 3）或用 *N*-乙基马来酰亚胺处理（泳道 4）失活了。泳道 5，用 BSA 代替组蛋白，与泳道 6（无蛋白质）一样，未显示活性。泳道 M 为蛋白质分子质量标准。（*Source*：Brownell，J. E. and C. D. Allis，An activity gel assay defects a single，catalytically active histone acetyltransferase subunit in *Tetrahymena* macronuclei. *Proceedings of the National Academy of Sciences USA*（July 1995）f. 1，p. 6365. Copyright © National Academy of Sciences. USA.）

Allis 及同事以这种初始的 HAT 活性鉴定为起点，应用经典分子克隆策略进一步研究了 p55 及其基因。他们用标准生化技术进一步纯化 HAT 活性。当纯化的 HAT 活性达到同质程度时，分离足够的量以检测部分氨基酸的序列。根据这个氨基酸序列，他们设计了编码部分氨基酸序列的一套简并寡核苷酸（第 4 章），从而与大核基因组 DNA 或细胞 RNA 进行杂交。用这些寡核苷酸为引物，细胞总 RNA 为模板，按第 4 章所述的方法进行 RT-PCR，然后克隆 PCR 产物，获得部分 PCR 产物的碱基序列，检测并证实其内部序列也可编码已知的 HAT 氨基酸序列。没有一个 PCR 克隆含有全长 cDNA，所以，他们采用快速 cDNA 末端扩增技术（RACE，见第 5 章）从 5′和 3′两个方向进行延伸，最终获得了编码 p55 蛋白质全部 421 个氨基酸的 cDNA 克隆。

从 p55 cDNA 碱基序列推导出的氨基酸序列与酵母 Gcn5p 蛋白的序列很相似。Gcn5p 被鉴定为酸性转录激活因子（如 Gcn4p）的辅激活因子，因此

氨基酸序列的相似性表明 p55 和 Gcn5p 两者都是与基因激活相关的 HAT。为了证实 Gcn5p 有 HAT 活性，Allis 及同事在 *E. coli* 中表达了 Gcn5p 基因，然后把 Gcn5p 和 p55 都进行 SDS-PAGE 活性胶分析，两种蛋白质都显示了明显 HAT 活性。因而，至少一个 HAT（Gcn5p）有 HAT 活性和转录辅激活因子活性，在乙酰化组蛋白调控基因活性的过程中起直接作用。

值得注意的重要一点是 p55 和 Gcn5p 都属于细胞核的 A 型 HAT（HAT A's），很明显参与基因的调控。它们将核心组蛋白的富赖氨酸 N 端尾部乙酰化。完全乙酰化的组蛋白 H3 在赖氨酸 9、14、18 都有乙酰基，完全乙酰化的组蛋白 H4 在赖氨酸 5、8、12、16 都有乙酰基。组蛋白 H3 的赖氨酸 9、14 和组蛋白 H4 的赖氨酸 5、8、16 在活性染色质中乙酰化，在非活性染色质中去乙酰化。B 型 HAT（HAT B's）发现于胞质中，使新合成的组蛋白 H3 和 H4 乙酰化并正确装配到核小体中。由 B 型 HAT 所加上的乙酰基团随后在细胞核中被组蛋白脱酰酶（histone deacetylase）去除。所有已知的 A 型 HAT，包括 Gcn5p 和 p55，都有溴基区（bromodomain），而所有已知的 B 型 HAT 缺乏溴基区。溴基区允许蛋白质结合到乙酰化的赖氨酸上，这对 A 型 HAT 非常有用，A 型 HAT 必须识别部分乙酰化的组蛋白尾部并把乙酰基团加到其他赖氨酸残基上。但是溴基区对 B 型 HAT 没有意义，因为 B 型 HAT 只能识别新近合成的非乙酰化的核心组蛋白。

自从 Allis 研究小组发现 p55 以后，除 Gcn5p 外的好几个辅激活因子都被鉴定出具有 HAT A 活性，其中有 CBP/p300（第 12 章）和 $TAF_{II}250$（第 11 章）。这 3 种辅激活因子与激活因子协同增强转录。它们都包括 HAT A 活性的事实提示了转录增强的部分机制：通过结合在转录起始位点附近，使其邻近核小体中的核心组蛋白乙酰化，中和其部分正电荷，松弛组蛋白对 DNA（可能是附近核小体上的 DNA）的结合，使染色质重建，以便更易进入转录装置，进而激活转录。

令人感兴趣的是 $TAF_{II}250$ 有两个溴基区模块能识别两个相邻的乙酰化赖氨酸。例如，我们可以在非活化染色质的部分乙酰化核心组蛋白上找到乙酰化赖氨酸。因而，$TAF_{II}250$ 的另一个功能可能是在非活化染色质中识别部分乙酰化的核心组蛋白，引导其伴侣 TBP 和其他 TAF_{II} 进入这种染色质，启动活化过程。在本章后面我们将看到这一假说的证据。

小结 在细胞质和细胞核中均可发生组蛋白乙酰化。细胞质乙酰化由 HAT B 执行，并为核小体的构建准备组蛋白，这些乙酰基团随后在细胞核中被除去。核内的核心组蛋白 N 端乙酰化由 HAT A 进行，并与转录激活相关。许多辅激活因子有 HAT A 活性，能够松弛核小体与基因调控区的结合。核心组蛋白尾的乙酰化也吸引溴基区蛋白如 $TAF_{II}25$，该蛋白质对转录至关重要。

组蛋白去乙酰化

如果核心组蛋白的乙酰化是个转录激活事件，可以预测核心组蛋白的去乙酰化应该是个抑制事件。与此假说一致的是，低水平乙酰化核心组蛋白的染色质比正常染色质的转录活性低。图 13.24 是这种抑制作用的相关机制：已知的转录抑制物，如未结合配体的核受体，与辅阻遏物相互作用，后者再与组蛋白脱酰酶作用，脱酰酶将邻近核小体的核心组蛋白的碱性尾上的乙酰基团除去，使组蛋白更紧密地与 DNA 结合，稳定了核小体，并维持转录的抑制状态。虽然不及染色体的异染色质域，如末端或端粒所致的沉默那么严重，这种抑制也可被称为**沉默**（silencing）。研究得很清楚的辅阻遏物有 SIN3（酵母）、SIN3A 和 SIN3B（哺乳动物）和 **NCoR/SMRT**（哺乳动物）。NCoR 表示“核受体辅阻遏物”（nuclear receptor corepressor），SMRT 表示“类维生素 A 和甲状腺激素受体沉默介质”（silencing mediator for retinoid and thyroid hormone receptor）。这些蛋白质均与无配体的视黄酸受体（RAR-RXR）（一种异源二聚体核受体）相互作用。

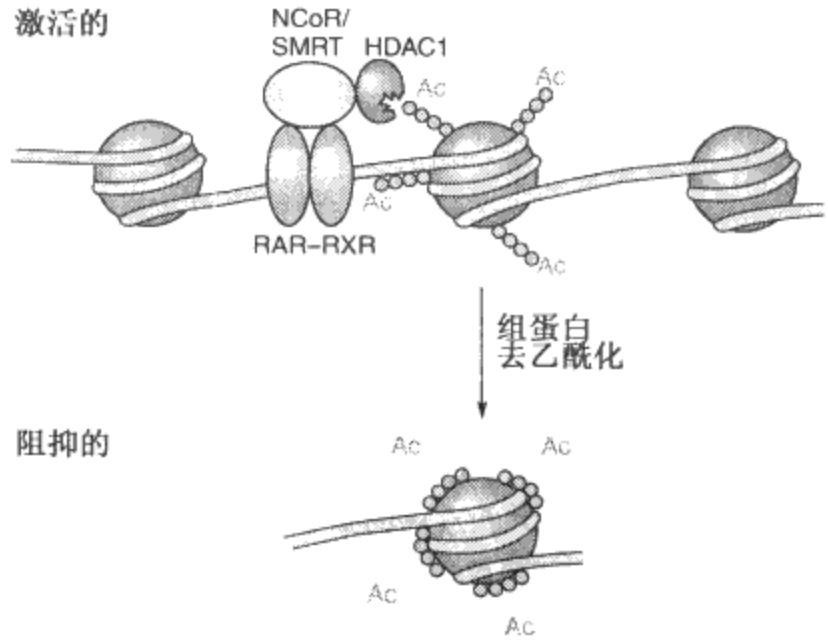


图 13.24 组蛋白脱酰酶参与转录抑制的模型。异源二聚体视黄酸受体（RAR）和视黄酸受体 X（RXR）结合到增强子上（顶端）。在无配体视黄酸时，受体二聚体结合辅抑制物 NcoR/SMRT，随之与组蛋白脱酰酶 HDAC1 结合。脱酰酶从邻近核小体核心组蛋白的赖氨酸侧链（灰色）去除乙酰基（红色）。这种脱酰基作用使赖氨酸侧链更紧密地与 DNA 结合（见底端），维持核小体稳定，导致转录抑制。

怎样才能知道存在于转录因子、辅阻遏物、组蛋白脱酰酶之间的物理联系呢？解决这个问题的方法之一是将抗原决定簇标签表位附加到其中一个组分上，然后用抗标签的抗体免疫共沉淀整个复合物。例如，Robert Eisenman 及同事用抗原决定簇标签证明了转录因子 **Mad-Max**、哺乳动物 Sin3 辅阻遏物（SIN3A）和组蛋白脱酰酶（HDAC2）可形成三元复合物。Max 是一种既可作为激活因子，也可作为抑制物的转录因子，取决于在异源二聚体中的搭档。如果它与 Myc 结合形成 Myc-Max 二聚体，就作为转录激活因子；反之，与 Mad 结合形成 Mad-Max 二聚体，就作为转录抑制物。

Mad-Max 引起抑制作用的部分原因是组蛋白去乙酰化作用，提示组蛋白脱酰酶与 Mad 存在某种相互作用。类比由图 13.24 所示的 RAR-RXR-NCoR/SMRT-HDAC1 的相互作用，我们推测某些辅阻遏物如 NCoR/SMRT 介导 Mad 与组蛋白脱酰酶间的相互作用。为表明这类作用确实在体内存在，并由辅阻遏物（SIN3A）介导，Eisenman 及同事采用了抗原决定簇标签策略。他们用两种质粒共转染哺乳动物细胞，一种质粒编码抗原决定簇标签的组蛋白脱酰酶［（FLAG- HDAC2）带小分子肽 FLAG 标签的 **HDAC2**］，另一种质粒是编码 Mad1 或脯氨酸发生替代的 Mad1 突变体（Mad1Pro），该脯氨酸既能阻断与 SIN3A 的相互作用又能阻断转录抑制作用。之后，从转染细胞中制备提取物，用抗 FLAG 抗体免疫沉淀复合物，然后经电泳、蛋白质转膜，先用抗 SIN3A 抗体探测，而后洗脱抗 SIN3A 抗体，再用抗 Mad1 抗体探测。

图 13.25 是实验结果。泳道 1～3 是阴性对照，细胞含有 FLAG-编码质粒，而不是 FLAG-HDAC2-编码质粒。用抗 FLAG 抗体对其细胞裂解物进行免疫共沉淀，不会有 HDAC2 或与之相关蛋白质的沉淀。因此，膜上找不到 SIN3A 和 Mad1。泳道 4～6 含有 FLDA-HDAC2-编码质粒和其他质粒共转染细胞的提取物，编码质粒是：无 Mad1（泳道 4）、Mad1（泳道 5）、Mad1Pro（泳道 6）。这 3 个泳道都含有 SIN3A，表明 SIN3A 与 FLAG-HDAC2 共沉淀了。只有泳道 5 含有 Mad1，泳道 4 没有 Mad1 是预期内的，因为没有 Mad1 质粒，泳道 6 很明显不含 Mad1Pro，即使在转染过程中添加编码该蛋白质的质粒。因为 Mad1Pro 不能结合 SIN3A，预期不会与 FLAG-HDAC2 产生共沉淀，除非它直接与 HDAC2 相互作用。事实上，没有共沉淀支持了这一假说。换句话说，辅阻遏物 SIN3A 介导转录因子 Mad1 和组蛋白脱酰酶 HDAC2 间的相互作用。泳道 7～9 显示没有任何免疫沉淀物的全细胞裂解物的简单电泳结果。这两张膜显示，细胞裂解物含有大量的 SIN3A 和丰富的 Mad1（如果有 Mad1-编码质粒）（泳道 8）或 Mad1Pro（如果该蛋白质存在）（泳道 9）。因此，泳道 6 没有 Mad1Pro 就不能用 Mad1Pro-编码质粒表达 Mad1Pro 蛋白失败来解释了。

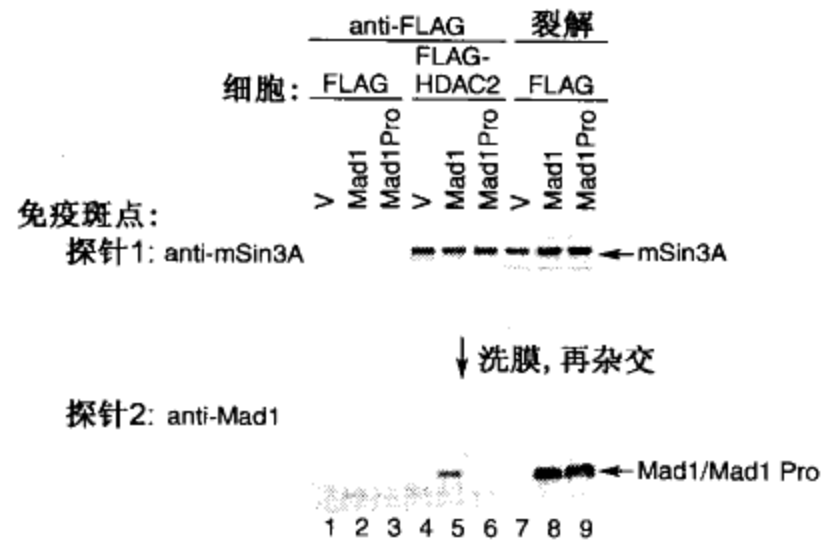


图 13.25　HDAC2、SIN3A、Mad1 三元复合物的证据。Eisenman 及同事用不同质粒转染细胞：只编码 FLAG 抗原决定簇或 FLAG-HDAC2，如图顶端“细胞”旁边的标注，或编码无 Mad1（V）、Mad1、Mad1Pro 也见图上所示。用抗 FLAG 抗体（泳道 1～6，图顶部标为“anti-FLAG”）免疫沉淀细胞裂解物，或只收集裂解物（泳道 7～9，图顶部标为“裂解”），电泳免疫沉淀物或裂解物。之后，将蛋白质转膜，先用抗 SIN3A 抗体探测（上方的印迹膜）。然后洗脱前一个膜，用既可与 Mad1 又可与 Mad1Pro 反应的抗 Mad1 抗体探测（下方的印迹膜）。最后，用连接有辣根过氧化物酶的二抗检测膜上的结合了抗体的蛋白质。再用与辣根过氧化物酶反应时变成化学发光物的底物检测该酶的存在。SIN3A 和 Mad1/Mad1Pro 的位置在印迹膜右侧标出。（*Source*：Laherty，C. D.，W.-M. Yang，J.-M. Sun，J. R. Davie，E. Seto，and R. N. Eisenman，Histone deacetylases associated with the mSin3 co-repressor mediate Mad transcriptional repression. *Cell* 89（2 May 1997）f. 3，p. 352. Reprinted by permission of Elsevier Science.）

现在我们已经看到两个蛋白质既是激活因子又是阻遏物的例子，这取决于与之结合的其他分子。有些核受体具有这个特性，取决于是否与配体结合。Max 蛋白的这个特性则取决于它是和 Myc 蛋白结合还是和 Mad 蛋白结合。图 13.26 图解了细胞核受体的这种现象，以**甲状腺激素受体**（thyroid hormone receptor，**TR**）为例。TR 和 RXR 形成异源二聚体并结合到增强子**甲状腺激素应答元件**（thyroid hormone response element，**TRE**）上。在无**甲状腺激素**

(thyroid hormone) 时，TR 是阻遏物。产生抑制的部分原因是它与 NcoR、SIN3 和组蛋白脱酰酶（如 mRPD3）相互作用。组蛋白脱酰酶能使核小体周围的核心组蛋白去乙酰化，从而稳定核小体，抑制转录。

在有甲状腺激素存在时，TR-RXR 二聚体成为激活因子。激活的部分原因是结合了 p300/ CBP、P/CAF、$TAF_{II}250$。这 3 个蛋白质都是组蛋白乙酰转移酶，能使邻近核小体中的组蛋白乙酰化，使核小体不稳定，从而激活转录。值得一提的是，组蛋白乙酰转移酶和组蛋白脱酰酶的功能靶标都是核心组蛋白，而不是组蛋白 H1。因此，核心组蛋白与组蛋白 H1 一样，在核小体稳定和非稳定状态中起重要作用。

核心组蛋白尾部的乙酰化作用显然并不仅仅是抑制这些尾部与 DNA 的结合。正如本章前述（图 13.4），Timothy Richmond 及同事的核心核小体颗粒的 X 射线晶体学揭示，在结晶网格中一个核小体核心颗粒的组蛋白 H4 与相邻核小体颗粒的组蛋白 H2A-H2B 二聚体产生相互作用。具体讲，组蛋白 H4 N 端的碱性区域（第 16～25 位氨基酸残基）与毗连核小体 H2A-H2B 二聚体的酸性口袋相互作用，这种相互作用有助于解释核小体间的交联，阻止接近转录因子，从而抑制转录。这个假说也有助于解释核心组蛋白尾部乙酰化为什么有激活转录的作用。乙酰化作用使组蛋白 H4 N 端尾部的正电荷被中和，有助于防止核小体交联，从而阻止转录的抑制。

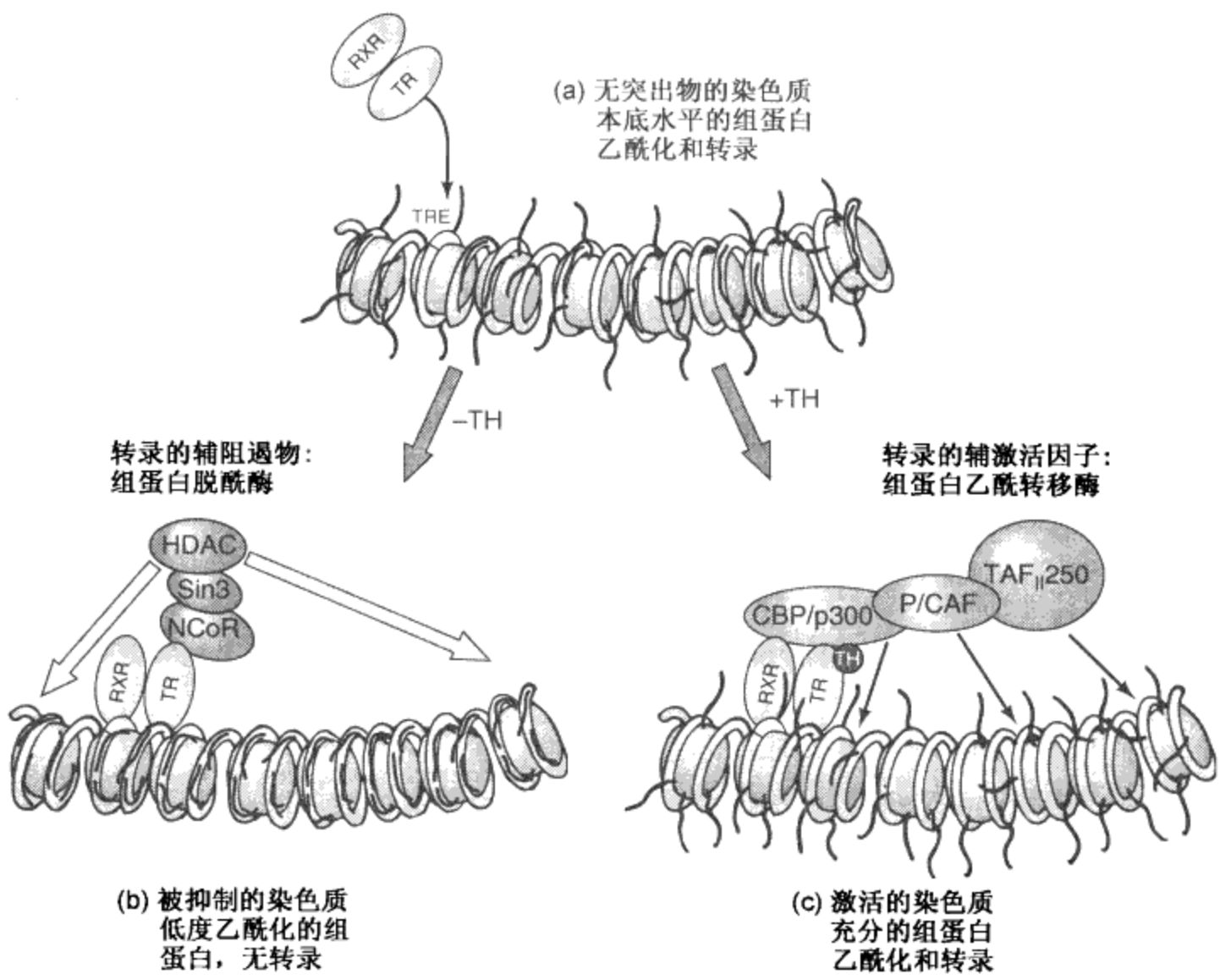


图 13.26　同一个细胞核受体的激活和抑制模型。（a）无突出物的染色质。无细胞核受体（TR-RXR 二聚体）结合甲状腺激素响应元件（TRE）。核心组蛋白尾略有乙酰化，转录在本底水平发生。(b) 被抑制的染色质。在无甲状腺激素（TH）时，核受体结合到 TRE 上，与辅阻遏物 SIN3 或 NCoR 相互作用，辅阻遏物又与组蛋白脱酰酶（HDAC）作用。脱乙酰基酶从核小体周围的核心组蛋白尾部切去乙酰基，使组蛋白与 DNA 的结合及周围核小体的组蛋白之间的结合更紧密，从而有助于抑制转录。(c) 活性染色质。甲状腺激素结合到核受体二聚体的 TR 部分，改变其构象以便与一个以上的辅激活因子 CBP/p300、P/CAF、$TAF_{II}250$ 结合。这些辅激活因子都是乙酰化周围核小体核心组蛋白尾的 HAT A，松弛组蛋白与 DNA 间的结合以及组蛋白与周围核小体的结合，有助于激活转录。(*Source*：Adapted from Wolfe，A. P. 1997. Sinful repression. *Nature* 387：16-17.)

然而，正如前一节所述，简单的电荷中和仅是此类事件的一部分。核心组蛋白尾部的乙酰化赖氨酸为溴基区蛋白（如 $TAF_{II}250$）提供了对接位点（docking site），溴基区蛋白对于转录至关重要。其实，我们在下一节将看到，核心组蛋白的乙酰化和其他修饰方式可构成“组蛋白密码”(histone code)，该密码能被其他激活和抑制转录的蛋白质所解读。

小结 转录抑制剂如无配体的核受体和 Mad-Max 结合到 DNA 位点，并与 NCoR/SMRT、SIN3 等辅阻遏物相互作用，进而结合组蛋白脱酰酶，如 HDAC1 和 HDAC2。这种三元蛋白复合物的装配使组蛋白脱酰酶更接近附近的核小体。核心组蛋白的去乙酰化使组蛋白的碱性尾部紧密结合到附近核小体的 DNA 和组蛋白上，稳定和交联核小体，从而抑制转录。核心组蛋白的去乙酰化作用也可去除转录激活必需的溴基区蛋白的结合位点。

染色质重建

组蛋白乙酰化作用常常是基因去抑制所必需但非充分的条件，因为乙酰化只是作用于核心组蛋白的尾部，而尾部位于核小体核心颗粒的外面。正如我们在下一节将看到的，组蛋白尾部的乙酰化可干扰核小体交联，留下了一个完整的核小体，因此另外还需要某种物质“重构”核小体核心颗粒，以允许其接近转录因子。这种重构过程需要 ATP 提供能量。

染色质重建复合物 依据 ATPase 的组成进行分类，至少有 4 类蛋白质参与**染色质重建**(chromatin remodeling)，ATPase 控制 ATP 水解，为染色质重建提供能量。4 类蛋白包括 **SWI/SNF** 家族（发音“switch-sniff”）、**ISWI** 家族（“limitation switch”）、**NuRD** 家族和 **INO80** 家族。所有 4 类蛋白质都能改变核小体核心结构，不仅使转录因子更易接近 DNA，而且还使核酸酶和其他蛋白质易于接近。

SWI/SNF 家族已经从酵母到人类的真核生物中分离到。最初是在酵母中识别的，发现它是调节 *HO* 内切核酸酶的基因，而该基因负责交配型转换(mating type switching)（因而得名称的“SWI”部分）。它们也调节 *SUC2* 基因，该基因编码启动蔗糖发酵过程的转化酶。因此，缺乏编码复合物中该亚基基因的突变体是蔗糖-非发酵体（sucrose non-fermenter，因而得名称的“SNF”部分）。SWI/SNF 家族都具有 BRG1 ATPase（或在某些生物中叫做 Brm）。Gerald Crabtree 及同事利用抗 BRG1 的抗体从几种哺乳动物中免疫沉淀出 SWI/SNF 复合物，发现了 9～12 个与 BRG1 共沉淀的 BRG1-相关因子(BRG1-associated factor，BAF)。

哺乳动物与酵母的 BAF 有许多相似性，但有些蛋白质却不同。此外，哺乳动物的 BAF 比其酵母对等物具有更大的多样性。这可能反映了哺乳动物相对于酵母在发育上的复杂性，不同的哺乳动物复合物参与发育的不同阶段。

有一种 BAF 叫做 BAF155 或 BAF170，取决于物种。它含有 SANT 结构域（“SANT”是首字母缩拼词，指在其中发现该结构域的 4 种蛋白质）。该域有一个三维结构，这个三维结构与转录因子 Myb 的 DNA 结合域（DNA-binding domain，DBD）相似，但是 SANT 与 Myb 的 DBD 之间某些氨基酸有差异，它不能结合 DNA。推测该域的 DNA 结合域由酸性氨基酸残基线性排列而成，这与它能结合碱性组蛋白，而不能结合酸性 DNA 的作用相一致。

染色质重建蛋白的 ISWI 成员也含有 SANT 结构域，实际上含有两个，一个是酸性氨基酸占优势的标准 SANT 域。另一个在中性 pH 下有一个净正电荷，可参与 DNA 结合，这个结构域叫做类 SANT 的 ISWI 结构域（SLIDE），以区分于普通的 SANT 结构域。SANT 和 SLIDE 结构域都是 ISWI 结合核小体以及其 ATPase 被核小体激活所需要的。因此，这两个结构域都允许 ISWI 对核糖体的结合、将激活信号传到 ISWI 的 ATPase 结构域，进而使染色质重建。

所有 4 类蛋白质都可在活跃基因的特征性增强子和启动子周围产生无核小体区。其实，我们推测无核小体区增强子是基因活化的重要前提。因此，毫不奇怪当许多酵母基因被激活时，SWI/SNF 是首先参与基因激活的辅激活因子之一。

小结 许多真核基因的激活需要染色质重建。几种不同的蛋白质复合物执行这一重构任务，它们都有 ATPase，可以控制 ATP 水解，为染色质重建提供能量。根据 ATPase 的组成划分重构复合物，其中两个研究得最清楚的复合物是 SWI/SNF 和 ISWI。哺乳动物的 SWI/SNF 复合物以 BRG1 为其 ATPase，有 9～12 个 BRG1-相关因子（BAF）。其中一个高度保守的 BAF 叫做 BAF155 或 BAF170，具有负责组蛋白结合的 SANT 结构域，可帮助 SWI/SNF 结合到核小体上。重构复合物的 ISWI 家族具有 SANT 结构域和另一个叫做 SLIDE 的结构域，SLIDE 结构域可能参与 DNA 结合。

染色质重建的机制 “重建”的准确含义仍不清晰。有时它涉及核小体从其起始位置移开，为转

录因子打开启动子。但重构并不是必须涉及核小体的简单滑动。例如，染色质重建可在核小体背靠背地排列穿过启动子的染色质中发生，而简单地滑动这些成串排列的核小体并不能打开足够量的 DNA。本章的后面会讨论，有时重构涉及松弛一个或多个核小体，以便另外的蛋白质，如 TFIID 将其移到一边去。目前对“重建”暂时的理解就是它能移动核小体，即核小体通过滑动或其他机制移动。这种移动可由重构复合物本身或其他蛋白质引起。

此外，染色质重建的功能并不总是激活转录，有时，所有已知重构复合物都会协同产生抑制作用。因此，核小体的重构使之更易从启动子移开而激活转录，但也会使核小体移到抑制转录的位置。实际上，NuRD 复合物的一个亚基是组蛋白脱酰酶，该酶有助于抑制转录。

Robert Kingston 及同事在 SWI/SNF 的 BRG1 亚基上研究了染色质重建活性的本质，他们推测重建一方面是使 DNA 更易接近，因此以 DNA 的可接近性作为重建活性的测定。他们设想了两个重构模型（图 13.27）。模型 1 涉及核小体 DNA 相对于组蛋白的几种构象形成。模型 2 涉及单一重建构象的形成，这种情况在 DNA 进入或退出核小体那一刻只是简单地从组蛋白剥离时会有发生，就像暴露在单个核小体中未催化的 DNA 的情况一样。模型 2 也适用于核小体只是简单地沿 DNA 滑动的情况，如同在体外加热的核小体中的情况一样。

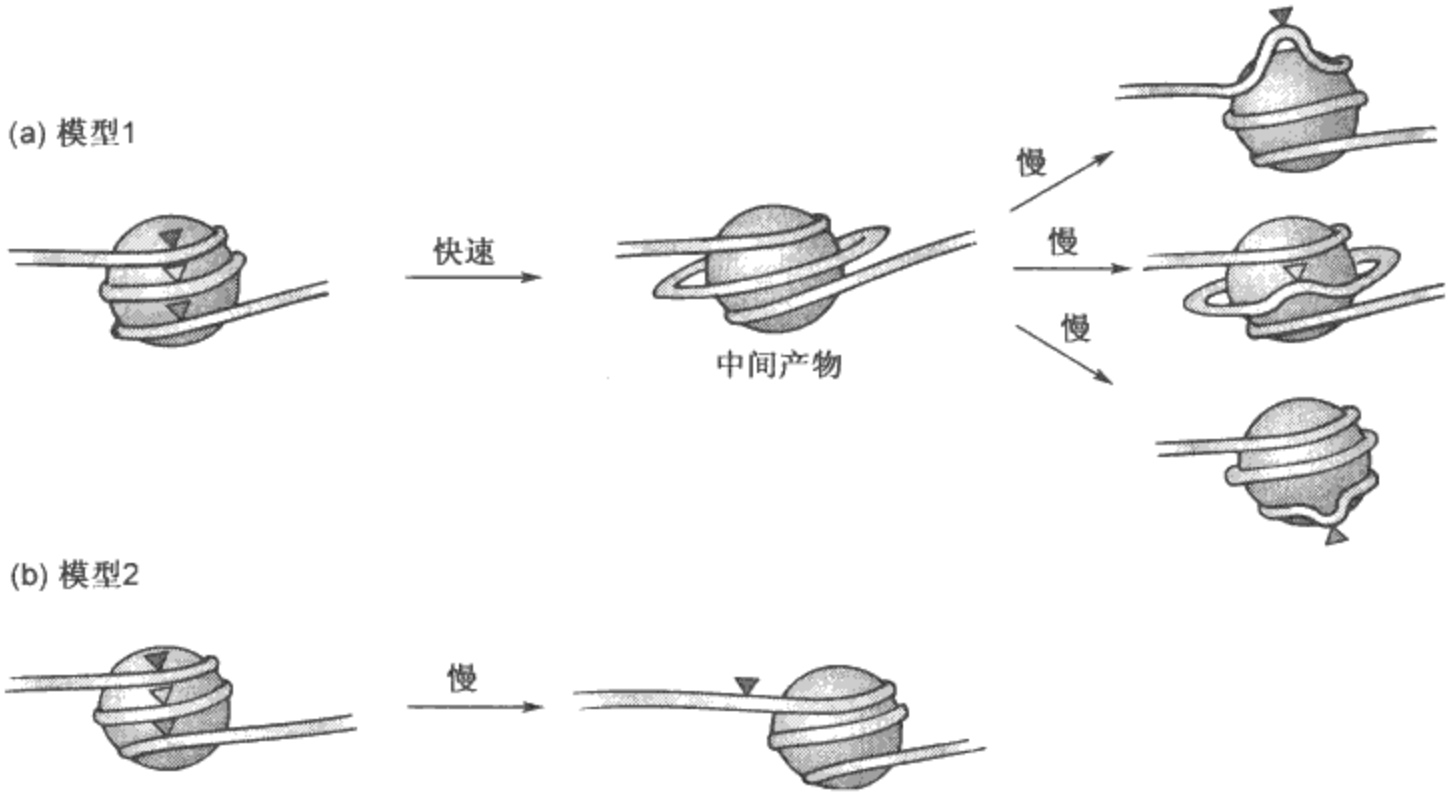


图 13.27　SWI/SNF 引起的染色质重建的两种模型。（a）模型 1，该核小体含有 3 个限制性位点，用彩色三角形表示。在第一步，核小体可能产生一个中间产物，该产物在限速步骤转换成各种重构构象。这里所示意的 3 种构象都敞开了一个限制性位点。(b) 模型 2，重构产生一个单一构象，在本例中敞开了其中一个限制位点。

Kingston 及同事设计了几种方法来区分这两种模型，结论都是模型 1 是正确的，重建染色质以几种不同的构象存在。他们以一种模式核小体开始，其中包含一个标记的 157bp 的 DNA 片段，该片段有 3 个限制酶 *Pst*I、*Spe*I、*Xho*I 的切割位点。他们推测两种模型在重建期间 3 个限制性位点的可利用速度不同。

注意，限制酶的实际切割速度非常快，所以不是限速步骤，而染色质构象变化能使限制位点的暴露变得相对较慢，是切割的限速步骤。因此，产生不同构象的模型 1 预测 3 种酶切割的速度会不同。因为不同的构象对 3 种酶具有不同的可接近性，并且不同构象形成的速度也不同。产生单一构象的模型 2 预期对 3 种酶具有相同的可接近性，因此应以相同的速度切割。

因此，Kingston 及同事将 BRG1 和 ATP 加入标记的模式核小体中，测定了重建期间每个限制酶的切割速度。图 13.28 显示切割速度有 9 倍的差异，支持了模型 1。此外，DNaseI 比 *Pst*I 的切割速度要快 10～20 倍，这一点也适合模型 1 但不适合模型 2。最后，Kingston 及同事用全 SWI/SNF 而不是 BRG1 重复了他们的实验，获得了相同的结果。因此，模型 1 也说明染色质重建也能由完整 SWI/SNF 发动，这些实验更清楚地证明真正催化的染色质重建明显不同于没有催化剂时所能发生的染色质简单改变。

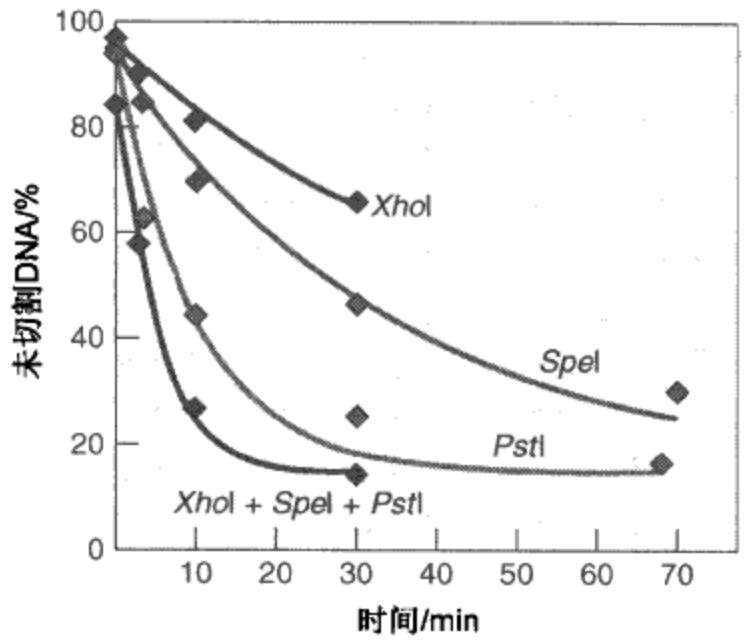


图 13.28　BRG1 催化的染色质重建过程中以不同速率所揭示的限制性位点。Kingston 及同事将 BRG1 标记的核小体与 ATP 培养不同时间至 70min，检测被 3 种限制酶 *Xho*I、*Spe*I、*Pst*I 切割的敏感性。去蛋白质 DNA 电泳确定未切割 DNA 的量，并以此对时间作图。(*Source*：Adapted from Narlikar G. J. et al.，*Molecular Cell* 8，2001. f. 4A，p. 1224.)

小结　对染色质重建机制的细节还不清楚，但涉及核小体的移动，伴随 DNA 与核心组蛋白间连接的松弛。与核小体中非催化 DNA 的暴露，或核小体沿一段 DNA 简单滑行不同，催化的核小体重建涉及核小体 DNA 关于核心组蛋白不同构象的形成。

在酵母 *HO* 基因激活中的重建　Kim Nasmyth 及同事研究了与酵母 *HO* 基因相关的蛋白质，该基因在酵母交配型转换中起关键作用。*HO* 基因的表达取决于细胞周期不同时期出现的一系列蛋白质。Nasmyth 及同事使用了一种叫做**染色质免疫沉淀**（chromatin immunoprecipitation，ChIP）的技术（图 13.29）。首先，将编码蛋白（Myc）的短区域（抗原决定簇）的 DNA 片段融合到编码已知 *HO* 基因相关蛋白的基因末端，产生 C 端有 Myc 抗原决定簇的融合蛋白。然后，使酵母细胞同步化，以便大多数细胞一起历经细胞周期。收获处于细胞周期不同时期的细胞，加入甲醛，使 DNA 和与之结合的任何蛋白质间形成共价键。之后，用超声波破碎染色质，产生短的与蛋白质偶联的双链 DNA 片段。接着，制备细胞提取物，用直接抗 Myc 抗原决定簇的抗体免疫沉淀蛋白质-DNA 复合物。免疫沉淀步骤同样利用连接到大珠子上的葡萄球菌蛋白 A 或蛋白 G。蛋白 A 或蛋白 G 结合到抗体上，大珠子使整个复合物在离心时容易沉淀下来。回想 Myc 抗原决定簇是黏附在已知的与 *HO* 基因结合的蛋白质上，所以，免疫沉淀的蛋白质-DNA 复合物应当包含融合蛋白和 *HO* 基因。为了证实这些复合物中含有 *HO* 基因，他们用 *HO*-特异性引物进行 PCR。如果 *HO* 基因确实存在，PCR 产物应为一条预期大小的条带。

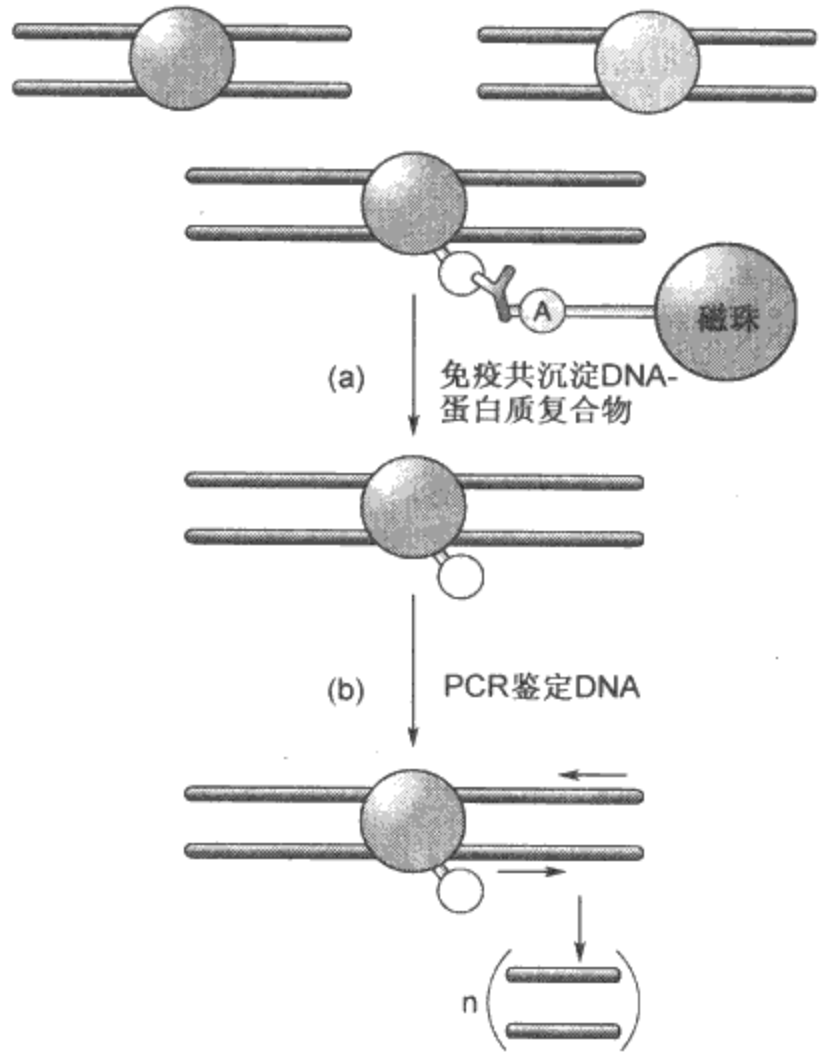


图 13.29　染色质的免疫沉淀。染色质与甲醛交联并剪成小片段。(a) 免疫沉淀步骤。抗体（红色）与吸附了目的蛋白（紫色）的抗原决定簇（黄色）结合，而目的蛋白又与双链 DNA 的特定位点（蓝色）结合。抗体结合葡萄球菌蛋白 A（或 G），后者又与可离心纯化的大珠偶联。大珠甚至可以是磁性的，用磁铁将免疫复合物拽到管子的底部。该抗体不与无抗原决定簇吸附的蛋白质（绿色和橙色）结合。(b) 免疫沉淀物中 DNA 的鉴定。用目的 DNA 特异引物进行 PCR，扩增部分 DNA。与预期大小相符的 DNA 产物表明该蛋白质确实与目的 DNA 结合（引物并不扩增蛋白质结合的精确 DNA 序列，而是目的基因邻近部分）。

实验结果显示，已知为 Swi5 的蛋白质首先与 *HO* 的调控区结合，接着是 SWI/SNF 结合，随后是含有 HAT Gcn5 的 SAGA 复合物结合（第 11 章），再召集激活因子 SBF。其他蛋白质，包括常见的转录因子和 RNA 聚合酶 II 在 SBF 之后依次结合。SWI/SNF 和 SAGA 绝对是 *HO* 激活所必需的，它们共同参与 *HO* 启动子周围染色质的重建。例如，SWI/SNF 能使该基因调控区周围的核心组蛋白解

体，SAGA 通过使核心组蛋白尾部乙酰化而增强这种解体作用，并有可能使这种解体永久化。其他研究有力地提示，这些因子并不是必须以上述提到的顺序发挥作用。在其他启动子上，它们能以多种不同的顺序发挥作用，并能相互协助完成其功能。在下一节，我们将看到基因在 SWI/SNF 复合物之前召集 HAT 的例子。

小结 ChIP 分析可以揭示在活化期间与基因结合的蛋白因子的顺序。当酵母 *HO* 基因活化时，首先结合的是 Swi5，接着是 SWI/SNF 和 SAGA，后者含有 HAT Gcn5p，然后是通用转录因子和其他蛋白。因此，染色质重建位于该基因活化的第一步，但是蛋白结合的顺序在其他基因上可以不同。

在人 IFN-β 基因中的重建：组蛋白密码 我们已经了解到核心组蛋白尾部的乙酰化和去乙酰化，它们分别趋于激活转录或抑制转录。但是组蛋白尾部还经受几种其他修饰，包括甲基化、磷酸化、泛素化和 SOUMO 修饰。其中每种修饰都影响周围基因的转录水平，由此产生了组蛋白密码（histone code）的概念。这一概念由 Thomas Jenuwein 和 David Allis 于 2001 年进行了详细描述，含义是在一个基因调控区附近的特定核小体上，组蛋白修饰的组合影响该基因的转录效率。组蛋白密码是一种**表观遗传**（epigenetic）密码（不影响 DNA 自身的碱基序列），是对基因及其调控区的碱基序列密码的补充。2001 年以来，许多研究都支持组蛋白密码假说。让我们来了解一下对人 IFN-β 基因的组蛋白密码研究。

Dimitris Thanos 及同事调查了在已透彻研究的人 IFN-β 基因激活期间发生的染色质重建过程。当该基因被病毒感染激活时，转录激活因子结合到启动子附近的无核小体区，形成增强体（第 12 章学过）。增强体中的激活因子召集一些因子，这些因子能修饰和重建转录起始位点周围的染色质，具体讲就是使一个核小体让开地方，以便启动转录。

这个过程涉及转录激活因子召集 HAT、SWI/SNF 复合物及一般转录因子等几个事件。HAT 使核小体的核心组蛋白尾部乙酰化，经 CBP 的一个或多个溴基区吸引 CBP-RNA 聚合酶 II 全酶。在全酶中 SWI/SNF 复合物松解核小体与启动子的结合。然后，TFIID 结合到 TATA 盒上并使之弯曲，此时，重建核小体向下游滑动到一个 36bp 的新位点，启动转录。

Thanos 及同事分析了核小体核心组蛋白的顺序乙酰化，发现组蛋白 H4 的 Lys8 乙酰化引起 SWI/SNF 复合物的召集，而组蛋白 H3 的 Lys9 和 14 乙酰化引起 TFIID 的召集。

他们利用 ChIP 分析技术观测仙台病毒（Sendai virus）感染 HeLa 细胞后，组蛋白乙酰化的时间进程，从此处着手，采用抗乙酰化和磷酸化组蛋白 H3 和 H4 的抗体免疫共沉淀交联的染色质。图 13.30a 显示，带有 IFN-β 基因的染色质可以被抗组蛋白 H4 的乙酰化 Lys8 和 Lys12 的抗体免疫沉淀，也可被抗组蛋白 H3 乙酰化 Lys9 和 Lys14 及磷酸化 Ser10 的抗体所免疫沉淀。但是，相同的染色质不能被抗组蛋白 H4 的乙酰化 Lys5 和 Lys16 的抗体沉淀。因此，组蛋白乙酰化的方式不是随机的。在另一个实验中，Thanos 及同事发现如果组蛋白 H4 的 Lys5 和 Lys16 确实被乙酰化，则抗该蛋白质的抗体就能沉淀该染色质。

此外，组蛋白修饰的时间进程会随着位置的不同而变化。病毒感染后 3～8h 后，组蛋白 H4 的 Lys8 被乙酰化，而其 Lys12 仅在感染 6h 后乙酰化。同样，组蛋白 H3 的 Ser10 磷酸化是在感染后约 3h 开始，6h 达高峰，其 Lys14 的乙酰化在感染约 6h 后开始，Lys9 的乙酰化开始得更早一些，但至少持续 19h。

组蛋白 H3 的 Ser10 磷酸化和 Lys14 乙酰化的时间进程支持一个早期的假说，该假说认为 Ser10 的磷酸化是 Lys14 乙酰化的必要条件。以上结果揭示了 Lys14 乙酰化进程与 TBP 对启动子结合的完全对应关系（比较第 9 排和第 10 排，显示抗 TBP 抗体的免疫沉淀）。这一结果与 H3 的 Lys14 乙酰化是 TBP 结合启动子所必需的假说一致。

Thanos 及同事在体外进行了相似的实验。他们用细菌表达的组蛋白重建了染色质并在体外修饰所选择的位点。结果发现体外乙酰化位点与体内乙酰化位点相同。而且，他们又用缺失一个或多个 HAT 的提取物进行相同实验，看 HAT 对特定赖氨酸乙酰化作用的影响。图 13.30b 显示 HAT GCN5/PCAF 的免疫耗竭提取物对组蛋白 H4 Lys8 的乙酰化没有作用。而 HAT CBP/p300 或 SWI/SNF 的成分 BRG1/BRM 的免疫耗竭提取物仍能乙酰化 H4 Lys8。另一对照实验证明 GCN5/PCAF 的耗竭不会引起 CBP/p300 的耗竭，反之亦然。因此，GCN5/PCAF 负责组蛋白 H4 Lys8 的乙酰化，而另一个实验（未给出），发现 HAT 负责组蛋白 H3 Lys14 的乙酰化（注意，GCN5 是酵母 Gcn5p 的人同源物）。

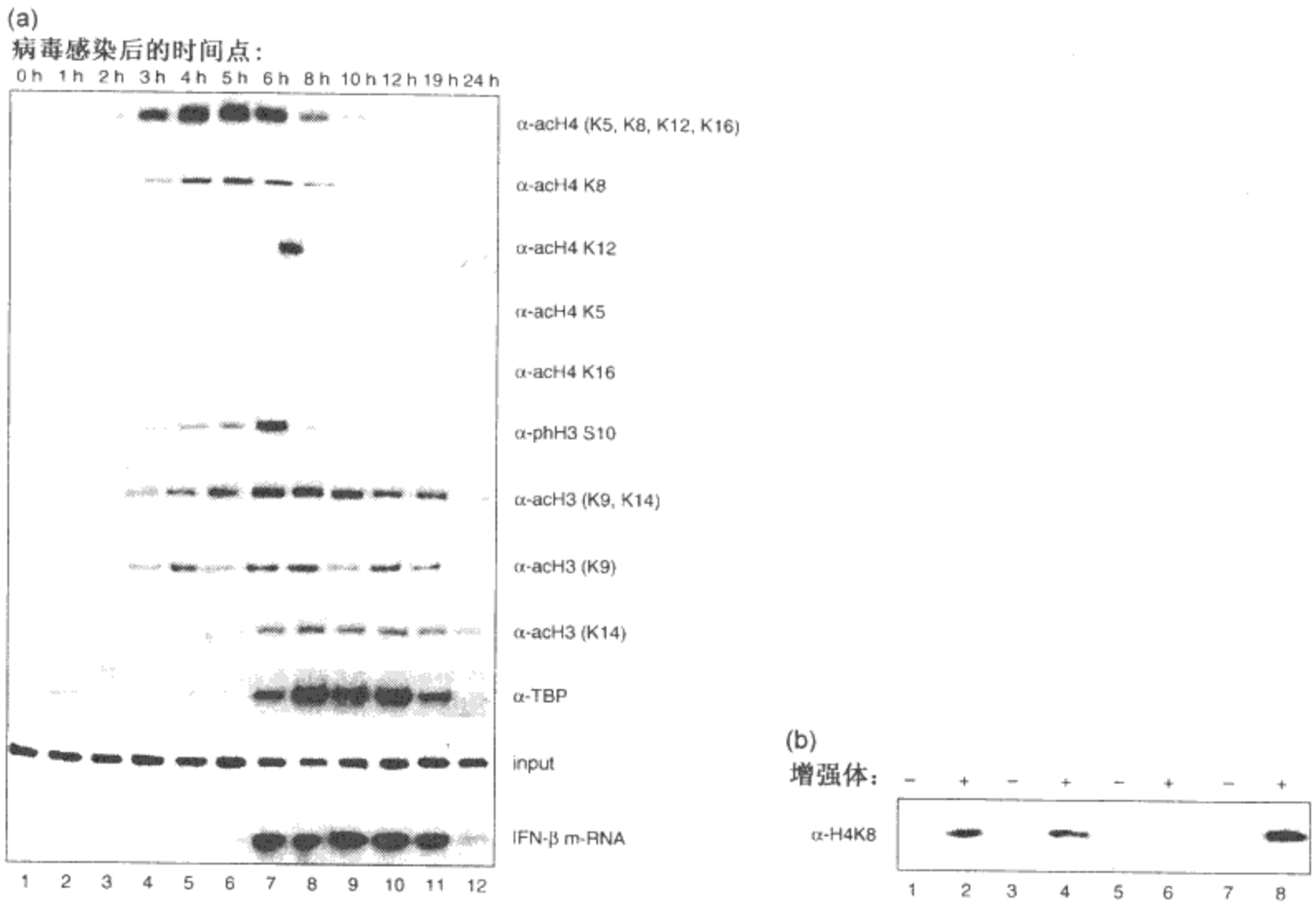


图 13.30　病毒感染后 IFN-β 启动子处染色质组蛋白的乙酰化进程。（a）ChIP 分析。Thanos 及同事对仙台病毒感染 HeLa 细胞后不同时间段的细胞核提取物进行了 ChIP 分析，所用抗体直接抗组蛋白 H4 的不同乙酰化位点（右侧标出）：Lys8（α-acH4 K8）、Lys12（α-acH4 K12）、Lys5（α-acH4 K5）、Lys16（α-acH4 K16）或所有这些抗体［α-acH4（K5、K8、K12、K16）］；或组蛋白 H3 磷酸化 Ser10（α-phH3 S10）；或组蛋白 H3 不同乙酰化位点：Lys9（α-acH3 K9）、Lys14（α-acH3 K14）、Lys9 和 14（α-acH3-K9，14）。他们做了抗体直接抗 TBP 的 ChIP 分析。之后，用 IFN-β 启动子特异引物对所有免疫沉淀染色质进行 PCR。图中给出了 PCR 结果，最底部是一组 RT-PCR 结果，表示 IFN-B mRNA 在不同时间点的丰度。input 行表示用加入的染色质获得的 PCR 信号，用以说明每个实验中所用染色质的量大致相等。(b) HAT 的免疫耗竭对组蛋白 H4 Lys8 乙酰化的影响。Thonas 及同事在生物素标记的 DNA 片段上装配了 IFN-β 增强体，这段 DNA 含有 IFN-β 启动子和增强子。然后将增强体（偶数泳道）或缓冲液（奇数泳道）与野生型细胞核提取物（泳道 1、2）或缺乏以下成分的细胞核提取物孵育：CBP/p300（泳道 3、4）、GCN5/PCAF（泳道 5，6）、SW1/SNF 的组分 BRG1/BRM（泳道 7、8）（译者注：原文遗漏）。然后，电泳蛋白质，Western 印迹，用直接抗组蛋白 H4 乙酰化 Lys8 的抗体探测印迹膜。(*Source*: Reprinted from *Cell* v. 111, Agalioti et al., p. 383. © 2002, with permission from Elsevier Science.)

为了研究核心组蛋白尾部乙酰化对 SWI/SNF 和 TFIID 召集的影响，Thanos 及同事用 IFN-β 启动子偶联到树脂小珠和核心组蛋白上构建了染色质，然后在乙酰基供体乙酰 CoA 有或无的条件下温育染色质与细胞核提取物，之后洗去未结合的蛋白质，再用 SDS 裂解染色质，对释放的蛋白质进行 Western 印迹，并用抗 BGR1（SWI/SNF 的一个成分）和 $TAF_{II}250$（TFIID 的一个成分）的抗体探测印迹膜。

图 13.31a 显示，染色质没有乙酰化时，只结合少量的 BRG1 和 $TAF_{II}250$（泳道 1 和泳道 2），而乙酰化时，可结合大量的 BRG1 和 $TAF_{II}250$（泳道 3 和泳道 4）。用 *E. coli* 的克隆基因产生的组蛋白来重建染色质，在无乙酰化时，检测不到 BRG1 和 $TAF_{II}250$（泳道 5 和泳道 6）的结合，但乙酰化后，能结合大量的 BRG1 和 $TAF_{II}250$（泳道 7 和泳道 8）。

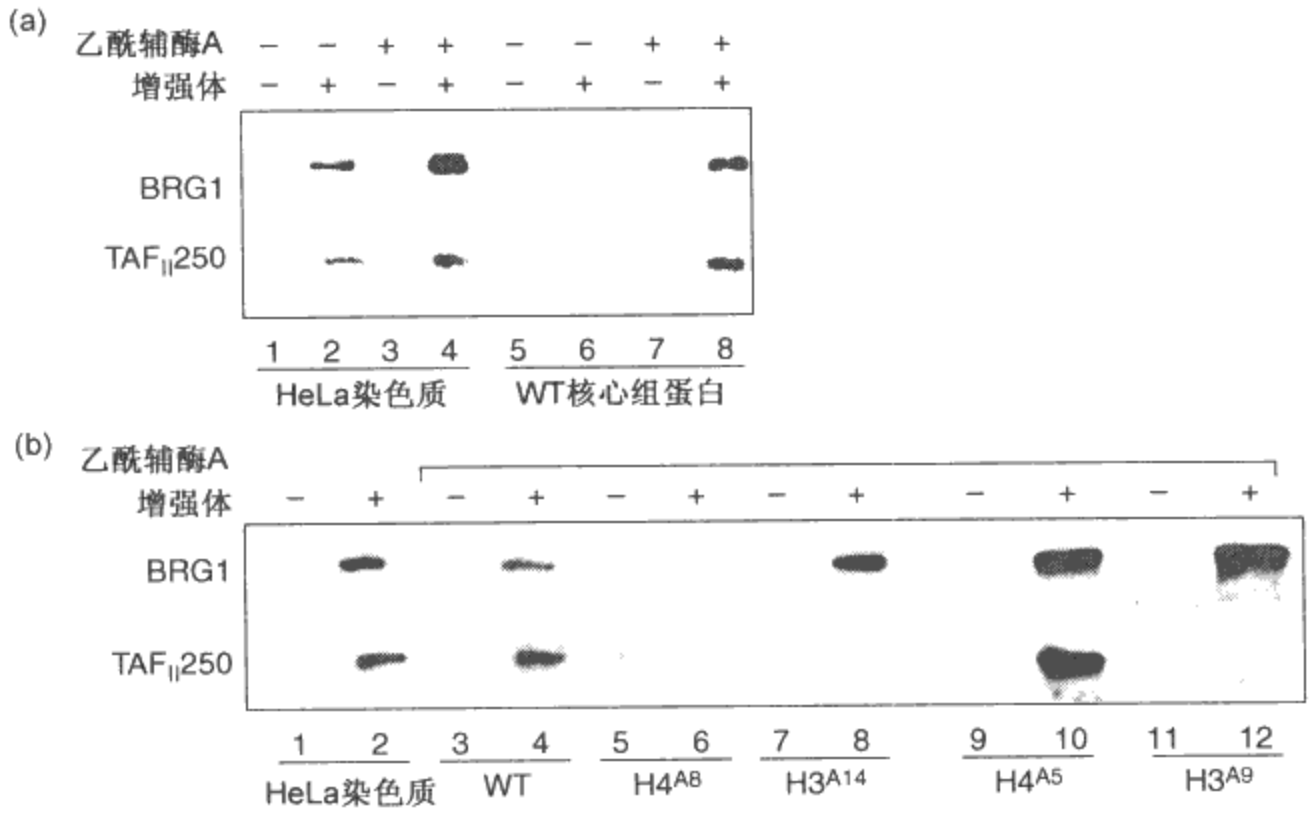


图 13.31　SWI/SNF 和 TFIID 向 IFN-β 启动子的汇集：在有野生型或突变核心组蛋白时。(a) Thanos 及同事在吸附到 Dyna-珠上的 IFN-β 启动子上重构了染色质，并与 HeLa 细胞核提取物共培养，洗去未结合蛋白质，Western 印迹检测 BRG1 和 TAF$_{II}$250 的结合，用抗这两种蛋白质的抗体探测。各泳道有无增强子，图上端标出在培养时加入乙酰 CoA 用于组蛋白乙酰化或不加乙酰 CoA，如图上端所示。泳道 1～4 为用天然 HeLa 细胞染色质重建的染色质。泳道 5～8 为 *E. coli*（WT 核心组蛋白）表达的重组野生型核心组蛋白构建的染色质。(b) 实验条件与 (a) 相同，除了在有些实验中采用突变核心组蛋白外（在泳道下面标出）。同样，有无增强体在顶端标出，同时有无乙酰 CoA 在括号中标出。举例突变体命名：H4^{A8} 表示组蛋白 H4 的赖氨酸 8 替换成丙氨酸。(*Source*: Reprinted from *Cell* v. 111, Agalioti et al., p. 386. © 2002. with permission from Elsevier.)

为了阐明特定组蛋白赖氨酸乙酰化的功能，Thanos 及同事用突变的组蛋白构建了染色质，其中一个赖氨酸被丙氨酸替换。图 13.31b 显示实验结果。如同在图 13.31a 看到的，天然 HeLa 细胞染色质对 BRG1 和 TAF$_{II}$250 都结合（泳道 1 和泳道 2）。可以预测，野生型组蛋白重建的染色质也可结合这两种蛋白质（泳道 3 和泳道 4）。但缺乏 Lys8（已被丙氨酸替代）的组蛋白 H4 重建的染色质不能和 BRG1 结合或不能与 TAF$_{II}$250 结合（泳道 5 和泳道 6）。这一结果可解释为突变体染色质不能召集 SWI/SNF（BRG1），而后者是召集 TFIID（TAF$_{II}$ 250）所必需的。

组蛋白 H3 Lys14 换成丙氨酸时，构建的染色质可召集 BRG1，但不召集 TAF$_{II}$250（泳道 7 和泳道 8）。同样的情况在组蛋白 H3 Lys9 换成丙氨酸时（泳道 11 和泳道 12）也能观察到。因此，Lys9 和 Lys14 的乙酰化看来是 TFIID 召集所必需的，但不是召集 SWI/SNF 所必需的。在对照实验中，将组蛋白 H4 Lys5 替换成丙氨酸，已知 Lys5 在病毒感染时不被乙酰化，因此该突变对 BRG1 或 TAF$_{II}$250 的召集都无影响并不奇怪（泳道 9 和泳道 10）。

采用同样方法，他们发现组蛋白 H3 Lys12 替换为丙氨酸对 BRG1 或 TAF$_{II}$250 的召集也不产生影响。该赖氨酸在体内可被乙酰化，但只是非常简短的（图 13.30）。显然这种乙酰化可能不是召集 TFIID 或 SWI/SNF 所需要的。最后，丙氨酸替换 Ser10 可阻断 TAF$_{II}$ 250 的召集但不影响 BRG1。因此，Ser10 缺失与 Lys9、14 缺失有相同效果。Ser10 缺失的效果与 Ser10 磷酸化是 Lys14 乙酰化所必需的这一假说一致。

所有这些结果可归纳为图 13.32 所示的模型。该模型的核心思想是增强子具有装配增强体所必需的所有遗传信息，然后增强体可召集适当的因子，移开阻断转录起始的核小体。因此，信息是从增强子向核小体流动的，而不是以相反的方向。

特别提出的是模型要求以下系列事件：病毒感染时，激活因子出现并在增强子上装配增强体。然后，增强体召集 HAT GCN5，后者又乙酰化组蛋白 H4 的 Lys8 和组蛋白 H3 的 Lys9。增强体也召集一种使组蛋白 H3 Ser10 磷酸化的未知蛋白激酶。一旦 Ser10 磷酸化，组蛋白 H3 的 Lys14 就被 GCN5 乙酰化。至此，组蛋白密码完成。

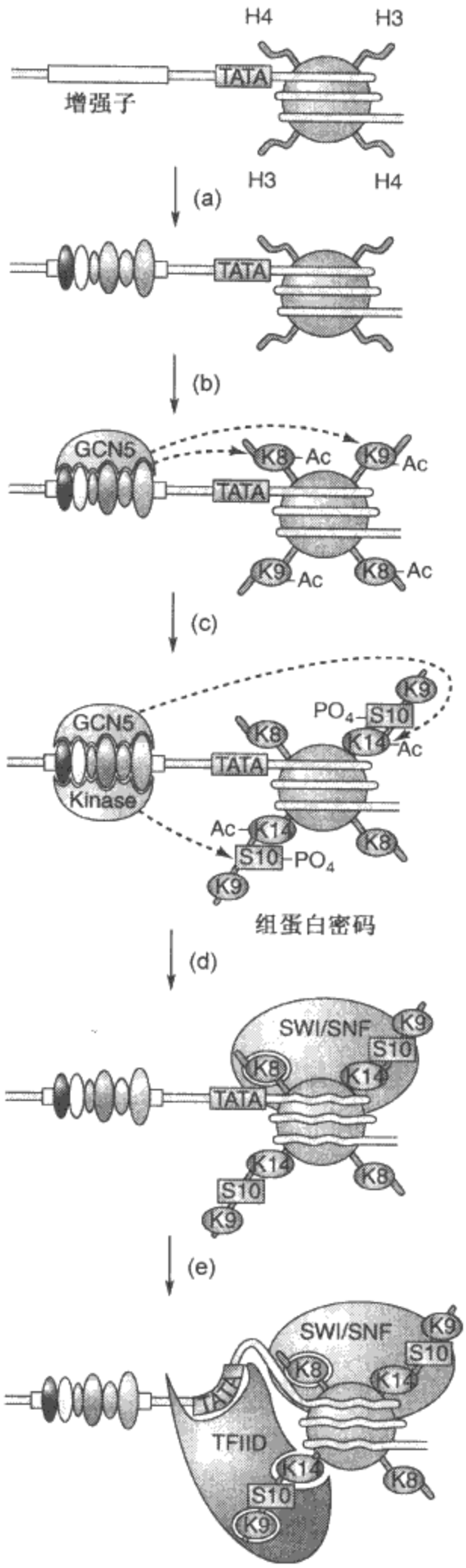


图 13.32　人类 IFN-β 启动子处组蛋白密码的模型。(a) 增强体按照 DNA 密码装配在启动子处(增强子元件的集合)。(b) 增强体中的激活因子召集 GCN5，这种 HAT 使组蛋白 H4 尾部的 Lys8（K8）和 H3 尾部的 Lys9（K9）乙酰化。箭头指出的乙酰化仅在较上部的组蛋白尾部，但乙酰化发生在所有 4 个尾部。(c) 增强体也召集使组蛋白 3 的 Ser10（S10）磷酸化的蛋白激酶。同样，磷酸化在 2 个 H3 尾部都发生。这种磷酸化作用允许 GCN5 对组蛋白 H3 的 Lys14（K14）乙酰化。由此完成组蛋白编码，其翻译在本模型的最后两步进行。(d) 组蛋白 H4 的乙酰化 Lys8 吸引 SWI/SNF 复合物，该复合物重建核小体。重建以核小体上弯曲的 DNA 表示。(e) 重建的核小体现在允许 TFIID 的结合了，TFIID 不仅被 TATA 盒吸引，也被组蛋白 H3 尾部的乙酰化 Lys9 和 14 所吸引。TFIID 使 DNA 弯曲，将重建的核小体推向下游 36bp。现在，转录可以开始了。(*Source*: Adapted from Agalioti, T., G. Chen, and D. Thanos, Deciphering the transcriptional histone acetylation code for a human gene. *Cell* 111 (2002) p. 389, f. 5.)

其次，含溴基区蛋白翻译组蛋白密码的方式为：单溴基区蛋白 BRG1 携带完整的 SWI/SNF 复合物与组蛋白 H4 乙酰化的 Lys8 结合，聚合酶 II 全酶的其余部分此时可能也被召集（为简单起见，未显示）。然后，SWI/SNF 重建核小体，双溴基区蛋白 $TAF_{II}250$ 有 2 个乙酰化赖氨酸（9 和 14），它携带完整的 TFIID 结合到组蛋白 H3 上。TFIID 的结合使 DNA 弯曲，使重建的核小体离开下游。现在，复合物可以与辅激活因子 CBP 结合了，转录开始。

这里值得一提的是 $TAF_{II}250$ 的另一个活性，它具有激活转录的潜力，虽然可能不像在 IFN-β 启动子处那样作用。也就是说，$TAF_{II}250$ 有泛素缀合(conjugating) 活性，其靶标之一可能是组蛋白 H1。因此，当 $TAF_{II}250$ 被召集到启动子之后，可能通过

结合到乙酰化核心组蛋白尾部，使附近的组蛋白 H1 泛素化，将其靶定，并由 26S 蛋白酶体降解（第 12 章）。由于组蛋白 H1 通过交联核小体而帮助抑制转录，所以，其降解会激活周围的基因。

小结 IFN-β 增强体的激活因子能召集一种 HAT（GCN5），GCN5 可使启动子处核小体的组蛋白 H3、H4 的一些赖氨酸乙酰化。一种蛋白激酶也能使同一核小体的组蛋白 H3 的一个丝氨酸磷酸化，由此允许组蛋白 H3 的另一个赖氨酸乙酰化，完成组蛋白编码。其中一个乙酰化赖氨酸再召集 SWI/SNF 复合物，重建核小体。这种重建使 TFIID 通过 $TAF_{II}250$ 的双溴基区与核小体上的两个乙酰化赖氨酸结合。TFIID 的结合使 DNA 弯曲，将重建核小体移开，为起始转录开通道路。

异染色质与沉默

本章中我们讨论的大部分染色质属于**常染色质**（euchromatin）。这种染色质结构相对伸展开放，并至少有潜在活性。相比而言，**异染色质**（heterochromatin）非常致密，其 DNA 不可接近。在高等真核生物中，用显微镜观察时甚至呈块状（图 13.33）。在酿酒酵母中，尽管染色体太小不能产生块状结构，但仍存在异染色质，并与高等真核生物中的一样具有抑制性特征，可沉默远至 3kb 外基因的活性。酵母的异染色质在染色体的端粒或末端都有发现，存在于永久被抑制的交配位点（mating loci）内（第 10 章结尾处提到的）。一般来讲，异染色质存在于染色体的端粒和着丝粒中。

图 13.33 间期核质中的异染色质。细胞核位于中心的蝙蝠胃的衬细胞。细胞核周边的深色区为异染色质（H）。(*Source*: Courtesy Dr. Keith Porter.)

在酵母中进行遗传和生化实验很方便，因此生物学家利用酵母来研究异染色质的结构及其沉默内部及邻近区段内基因的机制。由于基因沉默与该基因在染色体上的位置有关，因此端粒附近的基因沉默称作**端粒位置效应**（telomere position effect，**TPE**）：如果基因距端粒 3kb 以内，就会被沉默，如果超过此范围，则不被沉默。

对酵母端粒异染色质的研究表明，有多种蛋白质与端粒结合，推测可能与异染色质的形成有关。这些蛋白质有 RAP1、SIR2、SIR3、SIR4 以及组蛋白 H3 和 H4［SIR 表示沉默信息调节物（silencing information regulator）］。酵母端粒由许多 $C_{2\sim3}A(CA)_{1\sim5}$ 重复序列构成（显然，端粒的互补链具有互补序列）。通常被称为 $C_{1\sim3}A$ 的序列是 RAP1 蛋白的结合位点。RAP1 蛋白是唯一与 DNA 特异性位点结合的端粒蛋白。随后 RAP1 按 SIR3-SIR4-SIR2 的顺序召集 SIR 蛋白至端粒。我们已经知道，组蛋白 H3 和 H4 是核小体的核心组蛋白。SIR3 和 SIR4 直接与组蛋白 H3 N 端尾部的第 4～20 位氨基酸残基和 H4 N 端尾部的第 16～29 位氨基酸残基结合。

由于 RAP1 仅与端粒 DNA 结合，因此预计它仅与端粒有关。但是，在与端粒相邻的“亚端粒”（subtelomeric）区发现了 RAP1，且与 SIR 蛋白结合。为解释这一发现，Michael Grunstein 及同事提出了与图 13.34 相似的模型：RAP1 结合到端粒 DNA 上，SIR 蛋白与 RAP1 结合再与亚端粒区核小体的组蛋白结合，然后这种蛋白质-蛋白质相互作用引起端粒向亚端粒区回折。

在前面的章节，我们了解到除去核心组蛋白的乙酰基会抑制基因活性。因此推测，沉默染色质中核心组蛋白上的乙酰基不足，或乙酰基过低（hypoacetylated）。确实，在酵母常染色质中，组蛋白 H4 的 5、8、12、16 位赖氨酸均发生乙酰化，而酵母异染色质仅在 Lys 12 发生乙酰化。这种乙酰化过低现象在基因沉默中起什么作用呢？我们知道，组蛋白 H4 Lys16 是与 SIR 蛋白（特别是 SIR3）相互作用结构域的一部分（第 16～29 位氨基酸残基）。因此，组蛋白 H4 的 Lys16 乙酰化可能阻止与 SIR3 的相互作用，阻止异染色质的形成，从而阻止基因沉默。

对酵母的遗传学实验支持这一假说。将组蛋白 H4 的 Lys16 换成谷氨酰胺，通过去除正电荷来模拟该残基的乙酰化。这一突变可模拟通过乙酰化阻止基因（置于端粒邻近区域和交配基因位点处）沉默的作用。此外，如果将 Lys16 换成精氨酸（仍保持

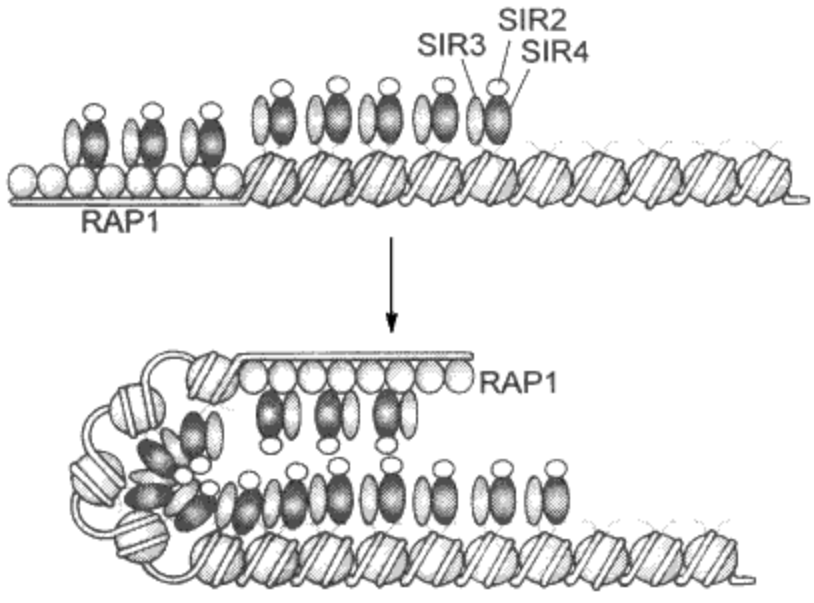


图 13.34　端粒结构的模型。RAP1 结合到端粒上并召集 SIR3 和 SIR4，它们再吸引 SIR2。SIR3、SIR4 还分别与组蛋白 H3、H4（蓝色细线）的 N 端尾部结合。SIR 蛋白质间的相互作用导致染色体末端的回折，因此 RAP1 与染色体亚端粒部分相结合。（*Source*：Adapted from Grunstein, M. 1998. Yeast heterochromatin: Regulation of its assembly and inheritance by histones. *Cell* 93: 325-28. Cell Press, Cambridge, MA.）

氨基酸的正电荷），结果在一定程度上模拟了去乙酰化，正如预测的，这种突变对基因沉默没有影响。

由于组蛋白 H4 的 Lys16 去乙酰化需要沉默复合物的参与，故发现了酵母沉默复合物的 SIR2 具有组蛋白脱酰酶活性（NAD 依赖的 HDAC，又称为 N-HDAC）令人十分感兴趣。因此，SIR2 是催化组蛋白 H4 的 Lys16 去乙酰化的候选酶。如果这个假说成立，则吸附在核小体（其组蛋白 H4 的 Lys16 去乙酰化）上的 SIR2 随后使邻近核小体组蛋白 H4 的 Lys16 去乙酰化，从而促进沉默过程。

小结　常染色质相对伸展、具有潜在活性，而异染色质致密、遗传学上处于不活跃状态。异染色质能使 3kb 以内的基因均发生沉默。酵母染色体末端（端粒）的异染色质形成依赖于蛋白 RAP1 对端粒 DNA 的结合，随后是对蛋白 SIR3、SIR4、SIR2 的依次召集。染色体其他位置上的异染色质形成也依赖蛋白 SIR。SIR3 和 SIR4 也直接与核小体组蛋白 H3、H4 相互作用。核小体组蛋白 H4 的 Lys16 乙酰化阻止其与 SIR3 的相互作用，因而抑制异染色质的形成。这是组蛋白乙酰化促进基因活性的另一种方式。

组蛋白的甲基化　除了我们已经看到的修饰方式，核心组蛋白尾部也可甲基化。甲基化有激活或抑制两种效应。某些蛋白质如 HAT，通过乙酰基-赖氨酸-结合域，也就是溴基区（bromodomain），与核心组蛋白尾上的特定乙酰化赖氨酸相互作用。Thomas Jenuwein 及同事注意到某些涉及异染色质形成的蛋白质具有称为着色域（chromodomain）的保守区域。其中一个具有这种结构的蛋白质是**组蛋白甲基转移酶**（histone methyl transferase，**HMTase**），在人类中被称为 SUV39H HMTase，另一个是组蛋白甲基转移酶-结合蛋白 **HP1**。

Jenuwein 小组和 Tony Kouzarides 小组验证了以上蛋白及其他蛋白对包含组蛋白 H3 Lys9（甲基化靶位点）的甲基化和非甲基化肽段的结合。两个小组都发现 HP1 可结合这些肽链，但只有 Lys 9 是甲基化的。这一发现提出了一个因为甲基化的扩展从而抑制染色质的机制：当一个组蛋白 H3 的 Lys9 甲基化后，会通过 Lys9 的着色域吸引 HP1，HP1 随后召集 SUV39H HMTase 使另一个邻近组蛋白 H3 的 Lys9 甲基化。以上过程持续进行直到许多核小体被甲基化。这种甲基化导致了异染色质区的扩展，如图 13.35 所示。

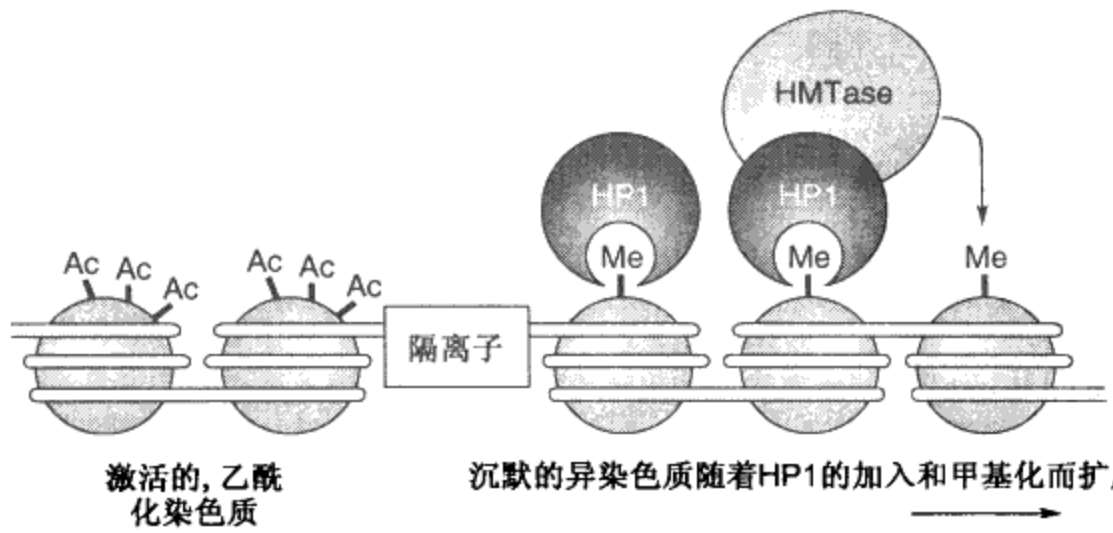


图 13.35　组蛋白甲基化参与染色质抑制作用的模型。隔离子右侧核小体组蛋白 H3 尾部的 Lys9 被甲基化。由此召集 HP1，它结合核小体的甲基化 Lys9，并召集组蛋白甲基转移酶（HMTase），使邻近核小体的 Lys9 甲基化。如此，甲基化抑制状态就从一个核小体传递至下一个核小体。（*Source*：Adapted from Bannister, S. D., P. Zegerman, J. F. Partridge, E. A. Miska, J. O. Thomas, R. C. Allshire, and T. Kouzarides, Selected recognation of methylated lysine 9 on histone H3 by the HP1 chromodomain. *Nature* 410 (2001) p. 123, f. 5.）

组蛋白 H3 的 Lys9 绝不是唯一的组蛋白甲基化靶位点。所有核心组蛋白的赖氨酸和精氨酸都能被甲基化，而且赖氨酸的每个氨基可接受多至 3 个甲基基团。组蛋白 H3 的另一个常被甲基化位点是 Lys4，该位点甲基化通常可激活转录，其机制至少涉及两个方面。第一，它抑制染色质重建和组蛋白脱酰酶复合物 NuRD 结合到组蛋白 H3 尾部，从而干扰了组蛋白去乙酰化作用，而该作用具有抑制转录的效果。第二，组蛋白 H3 Lys4 的甲基化抑制了相邻 Lys9 的甲基化，而 Lys9 甲基化也是转录抑制性的。通过阻止以上两个抑制事件，H3 的 Lys4 甲基化产生一个网状激活效应。

至此，我们已经孤立地研究了甲基化作用，但实际上并不以这种方式发生。相反，特定核小体的许多组蛋白残基有多种修饰方式。有的被乙酰化，有的被甲基化，还有的被磷酸化，甚至有的被泛素化。图 13.36 总结了核心组蛋白可能发生的不同修饰方式。

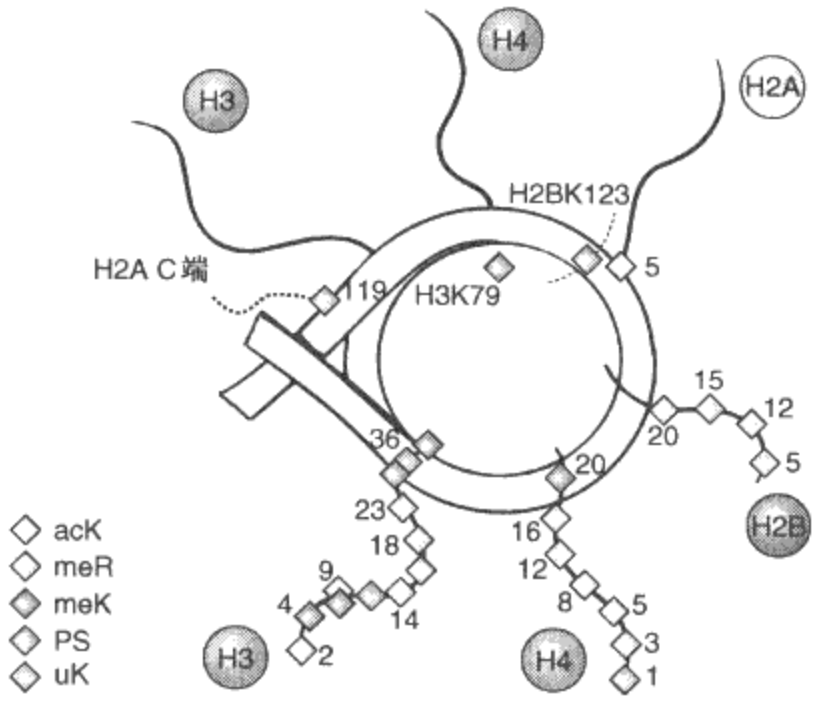


图 13.36　核心组蛋白修饰方式总结。不同修饰方式如左下端标识：黄色，乙酰化赖氨酸（acK）；灰色，甲基化精氨酸（meR）；蓝色，甲基化赖氨酸（meK）；粉色，磷酸化的丝氨酸（PS）；绿色，泛素化赖氨酸（uK）。仅在两种组蛋白 H3 和 H4 一侧的尾部显示了修饰方式。H2A、H2B 仅显示被修饰的一条尾。H2A、H2B 的 C 端尾部用虚线表示。组蛋白 H3 Lys79（H3K79）的位置也在图中标出，虽然它不在组蛋白尾部。（*Source*：Adapted from Turner, B. M.，Cellular memory and the histone code. *Cell* 111（2002）p. 286，f. 1.）

正如我们看到的，有证据表明组蛋白密码中组蛋白乙酰化参与了最终导致基因激活的级联事件。有些研究人员好奇，这种组蛋白密码的概念是否能类推到所有组蛋白修饰方式中。细胞能阅读特定核小体组蛋白修饰方式的不同组合，将其看做一个综合密码来指示附近基因的表达和沉默。

为了在组蛋白甲基化环境下说明这个问题，Frank Sauer 及同事研究了两种组蛋白中 3 个赖氨酸（组蛋白 H3 的 Lys4、Lys9 和组蛋白 H4 的 Lys20）的甲基化协同效应。他们发现由一个叫做 Ash1 的 HMTase 所产生的甲基化赖氨酸的组合对果蝇有两种效果，且都是激活功能。其一，这些甲基化可促进激活因子 Brahma 的结合；其二，这些甲基化阻止抑制物 HP1 和多聚梳状物（polycomb）的结合。因此组蛋白 H3 的 Lys9 甲基化的正常抑制作用在上述两个赖氨酸甲基化的背景下被掩盖。细胞一定能识别组蛋白修饰方式的整体组合，而不仅仅是其中的一个。

组蛋白修饰方式不仅标记了有待激活或抑制的染色质，也影响其他组蛋白的修饰。例如，组蛋白 H3 的 Lys9 甲基化可被同一组蛋白尾部的几种修饰作用所抑制，包括 Lys9（也可能是 Lys14）的乙酰化、Lys4 的甲基化、Ser10 的磷酸化。

同一核小体上一个组蛋白的修饰也能影响另一个组蛋白的修饰。例如，Brian Strahl 及同事检测了敲除酵母 *rad6* 基因的效果，该基因编码泛素连接酶 Rad6，而 Rad6 是组蛋白 H2B 的 Lys123 泛素化所必需的。这种突变阻断了 H3 的 Lys4、Lys79 的甲基化，但不影响 Lys36 甲基化（图 13.37）。用精氨酸替换组蛋白 H2B 的 Lys123 将阻止野生型 *rad6* 基因细胞的泛素化，并对组蛋白 H3 的 Lys4、Lys79 的甲基化有同样的负效应。因此，一种组蛋白（H2B）赖氨酸的泛素化会显著影响另一种组蛋白（H3）中两个位点的甲基化。顺便提一下，Lys79 的甲基化是唯一已知的不在组蛋白尾部发生的组蛋白修饰方式。该甲基化发生在核小体的表面（图 13.36），容易与甲基化装置接触。

最后，我们讨论一下组蛋白 H3 尾部 3 个氨基酸（Lys9、Ser10、Lys14）不同修饰之间的相互调节作用（图 13.38）。此前我们已经了解到，某些基因包括人类 IFN-β 基因的激活需要 Lys14 乙酰化，但是我们也知道这种乙酰化依赖于 Ser10 磷酸化，而 Ser10 磷酸化受 Lys9 甲基化的抑制。因此，Lys9 甲基化可以抑制 Ser10 磷酸化，进而阻断 Lys14 乙酰化，从而抑制转录。但事情的另一方面是，Ser10 磷酸化可能与 Lys14 乙酰化一起阻止 Lys9 的甲基化。因此，一旦 Ser10 和 Lys14 被适当地修饰，它们就可能通过防止抑制性的 Lys9 甲基化而维持基因的活化状态。此外，Lys9 乙酰化可阻止其甲基化，结果乙酰化也对转录抑制起到消减作用。

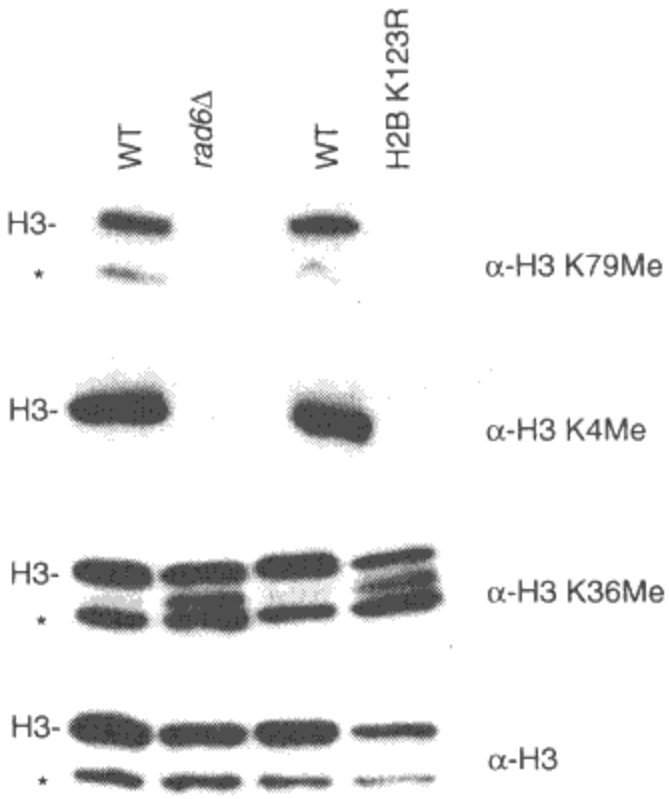


图 13.37　组蛋白 H2B 泛素化对组蛋白 H3 甲基化的影响。Strahl 及同事测试了野生型和突变酵母对组蛋白 H3 的 Lys79 甲基化的能力。突变株（*rad6Δ*）缺失了 *rad6* 基因，因而不能使组蛋白 H2B 的 Lys123 泛素化。突变株（H2B K123R）组蛋白 H2B 的 Lys123 被精氨酸取代，即使在具备 Rad6 功能情况下，也不能被泛素化。对野生型（泳道 1 和 3）、*rad6Δ*（泳道 2）、H2B K123（泳道 4）的细胞核提取物进行 Western 印迹，并分别用下列抗体做探针检测：Lys79 甲基化的组蛋白 H3（顶端一排）；Lys4 甲基化的组蛋白 H3（第二排）；Lys36 甲基化的组蛋白 H3（第三排）；组蛋白 H3（最下面一排）。最后一排以 H3 抗体为探针，作为阳性对照以确认每一泳道中均含有组蛋白 H3。突变株未表现出对组蛋白 H3 的 Lys4 或 Lys79 甲基化的支持，但确实支持 Lys36 的甲基化。星号代表除去 Lys4 甲基化位点的 H3 的蛋白水解产物。（*Source*：Reprinted with permission from *Nature* 418：from Briggs et al.，fig. 1，p. 498 © 2001 Macmillan Magazines Limited.）

小结　组蛋白 H3 N 端尾部的 Lys9 甲基化吸引蛋白质 HP1，随后 HP1 召集组蛋白甲基转移酶（假设可能使邻近核小体上的 Lys9 甲基化），从而扩展被抑制的异染色质化的状态。核心组蛋白中其他赖氨酸和精氨酸侧链的甲基化有抑制或激活双重效应，这些甲基化作用与其他组蛋白修饰方式，包括乙酰化、磷酸化、泛素化等同时发生在特定的核小体上。原则上说，每种特定的修饰组合能给细胞传递不同的转录抑制或激活的信息。一定的组蛋白修饰对其他邻近的修饰方式也产生影响。

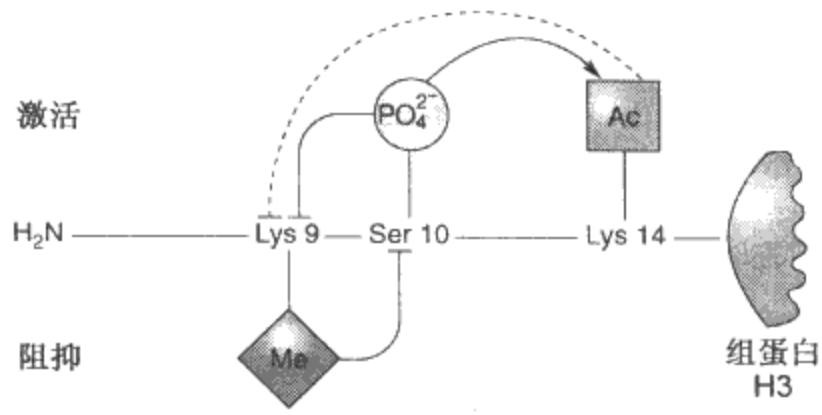


图 13.38　组蛋白 H3 N 端尾部不同修饰的相互作用模型。图中所示尾部上方的修饰起激活作用，包括 Ser10 磷酸化和 Lys14 乙酰化。尾部下方的修饰（Lys9 甲基化）起抑制作用。Ser10 磷酸化促进了 Lys14 乙酰化（箭头），但 Lys9 甲基化抑制了 Ser10 磷酸化（阻止符）。再者，Ser10 磷酸化可能还包括 Lys14 乙酰化，共同抑制了 Lys9 甲基化。

核小体与转录的延伸

我们知道当激活因子和通用转录因子与其相应的 DNA 位点结合时，基因的调控区无核小体，或至少核小体是被挤在一边的。但是 RNA 聚合酶如何处理结合在基因转录区域的核小体呢？

一个重要的因子是 **FACT** 蛋白［促进染色质转录（facilitate chromatin transcription）］。在体外，FACT 可促进 RNA 聚合酶 II 穿越核小体的转录延伸。人类 FACT 由两条多肽组成，一条是酵母 Spt16 蛋白的同源物，另一条是类 HMG-1 蛋白 SSRP1。FACT 与组蛋白 H2A 和 H2B 有很强的相互作用，由此提出假设认为 FACT 可以将这两个组蛋白从核小体移开，或至少暂时移开，从而使核小体不稳定以便 RNA 聚合酶可以转录通过。

有几个证据支持这一假设。首先，将 H2A 和 H2B 交联到核小体上无法离开，可阻断 FACT 的作用；第二，酵母编码组蛋白 H4 的基因突变，可改变组蛋白-组蛋白相互作用的方式，产生与 FACT 的 Spt6 亚基编码基因突变相同的表型。最后，活跃转录的染色质缺乏组蛋白 H2A 和 H2B。

2003 年，Danny Reinberg 及同事提供了 FACT 促进 RNA 聚合酶 II 转录的直接证据，即从核小体上至少移开一个 H2A-H2B 二聚体。他们也发现 FACT 具有**组蛋白分子伴侣**（histone chaperone）活性，可以将组蛋白送回染色质，在转录机构通过后重建核小体。

首先，研究人员利用免疫共沉淀实验显示 FACT 的 Spt16 亚基结合到 H2A-H2B 二聚体上，SSRP1 亚基结合到 H3-H4 四聚体上。Spt16 亚基具有强酸性 C 端，Spt16 亚基缺失 C 端的重组 FACT

(FACTΔ) 既不能与核小体组蛋白相互作用，也不能促进转录穿过染色质。

其次，用两种不同的荧光标签标记 H2A-H2B 二聚体和 H3-H4 二聚体。在 FACT 或 FACTΔ 处理之后，用含有 350mmol/L KCl 的缓冲液洗涤，然后进行 SDS-PAGE，测定二聚体/四聚体比值，荧光成像检测核小体上两种二聚体的丢失情况（荧光成像仪可定量地测定凝胶中条带的荧光强度）。图 13.39 显示 FACT 可引起处理过的核小体中 50% 的 H2A-H2B 二聚体丢失，而 FACTΔ 所引起的二聚体丢失不少于单独用缓冲液洗涤所引起的丢失（约 20%）。因此，看来 FACT 可减弱 H2A-H2B 二聚体与 H3-H4 二聚体之间的结合，这种影响依赖于 Spt16 亚基的 C 端。

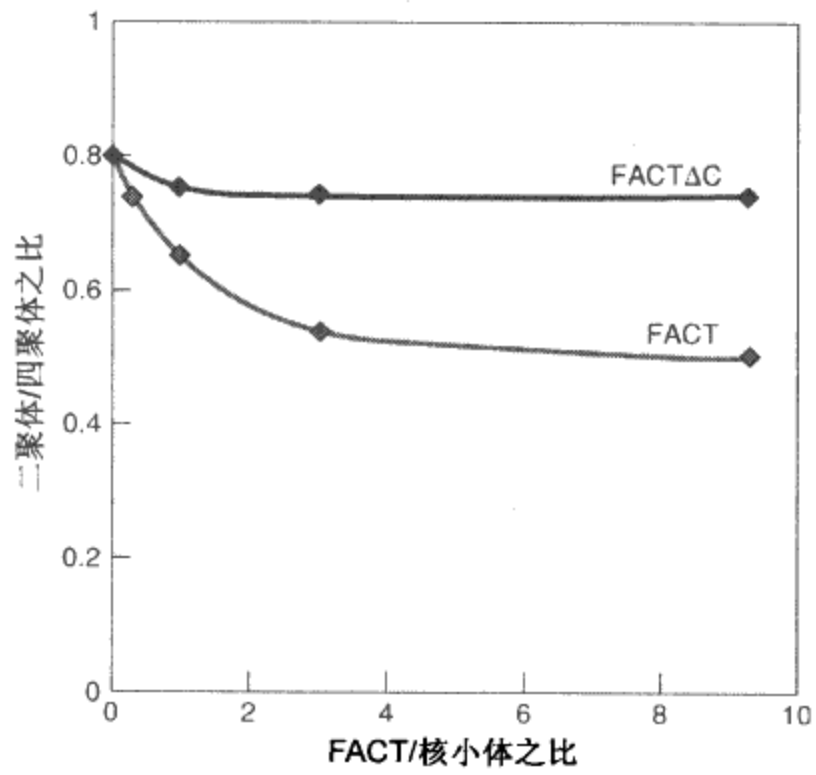


图 13.39　FACT 促进核小体上 H2A-H2B 二聚体的丢失。Reinberg 及同事用两种不同的荧光标签标记了 H2A-H2B 二聚体和 H3-H4 四聚体，加入 FACT 或 FACTΔ 培养 1h，洗涤核小体除去松散结合的组蛋白。然后通过 SDS-PAGE，测定二聚体/四聚体比值来检测核小体上两种二聚体的丢失情况，荧光标签在 SDS-PAGE 胶上用荧光成像仪定量测定。（*Source*：Adapted from Belotserkovskaya, et al., *Science* 301, 2003, f. 3, p. 1092.）

Reinberg 及同事也证明 FACT 可促进转录穿越核小体，并且转录模板含有**六聚体**（hexasome），这是一种缺乏 H2A-H2B 二聚体的核小体。为实施这一实验，他们以位于转录起点下游的单核小体为模板，将转录复合物组装到模板上，通过 RNA 聚合酶 II 上的标签把复合物连接到珠子上。然后在有 FACT 或 FACTΔ 情况下以标记的核苷酸进行转录。电泳检测转录物时，他们发现 FACT，可促进转录通过核小体形成全长的截断转录物，而 FACTΔ 却不能，即无 FACT 或有 FACTΔ 的转录产生若干转录物停滞在与核小体有关的 DNA 区，而有 FACT 的转录可减少这种停滞，产生更高比例的全长转录物。

该实验也能研究模板和全长截断转录物从 RNA 聚合酶上的释放。他们在转录前标记了 DNA，然后电泳检测释放的模板，模板应该是完全转录的。结果表明，有 FACT 的转录，模板含有六聚体，而有 FACTΔ 的转录没有六聚体。将 H2A 和 H2B 回加到六聚体中，可使它们转换成完整大小的核小体，表明六聚体确实是缺失了 H2A-H2B 二聚体的核小体。因此，FACT 至少是部分地将核小体结构松弛到足够的程度，至少允许丢失一个 H2A-H2B 二聚体来促进转录通过核小体。

FACT 仅使核小体解聚，它还可将组蛋白放回 DNA 重建核小体。Reinberg 及同事用两个实验证明了 FACT 的这种组蛋白分子伴侣作用。首先，将核心组蛋白与标记的 DNA 及无 FACT、FACT、FACTΔ 分别混合，电泳检测产物。结果表明，无 FACT 的转录混合物形成结块，不能进入电泳胶，而有 FACT 的转录可形成电泳行为良好的 DNA-组蛋白复合物。正如预测的一样，FACTΔ 不形成 DNA-组蛋白复合物。在第二个实验中，他们用两种不同的荧光标签标记 H2A-H2B 二聚体和 H3-H4 四聚体，然后用荧光成像仪在电泳凝胶上观察组蛋白-DNA 复合物是否含有与两组组蛋白结合的荧光标签。结果，FACT 具有组蛋白分子伴侣活性。他们也发现两个单独的 FACT 亚基都没有这一活性。

如果在转录延伸过程中 FACT 确实起到染色质重建者的作用，就应该在有 RNA 聚合酶的染色质上发现 FACT。Reinberg 和 John 及同事利用果蝇热激基因 *hsp70* 为实验体系，证明了这种行为。首先，他们分离了热激前后 20min 的果蝇多线染色体（polytene），与荧光标记的直接抗 RNA 聚合酶 II 和 Spt16 的抗体染色。结果表明热激之后，两个抗体共同定位在跨两个 *hsp*70 位点的染色体结块上。

如果 FACT 确实与 RNA 聚合酶 II 相伴，在转录进程中重构染色质，那么 FACT 应该像聚合酶 II 那样快速地被召集到热激基因上，应该在转录发生后立刻在结合转录因子-启动子的下游发现。为了检验这个假设，Reinberg 和 Lis 及同事查看了染色质与抗 FACT 两个亚基的抗体和抗 HSF（结合到 *hsp70* 基因上游调控区的激活因子）抗体染色的结果。

图 13.40 显示，热激 2.5min 之后，像 HSF 一

样，FACT 的两个亚基就与 *hsp70* 结合了。但是 FACT 的两个亚基都明显局限在比 HSF 更远的下游。通过比较 SSRP1 或 Spt16 染出的红色与 HSF 染出的绿色，可以看出这种差异。分开时很难看出，但是将两个图像叠加时，就可看出红色的前缘（FACT 荧光）在黄色的下游，黄色是红色（FACT）和绿色（HSF）荧光的重叠。这一效果在热激 10min 后也出现了，尤其是 SSRP1。

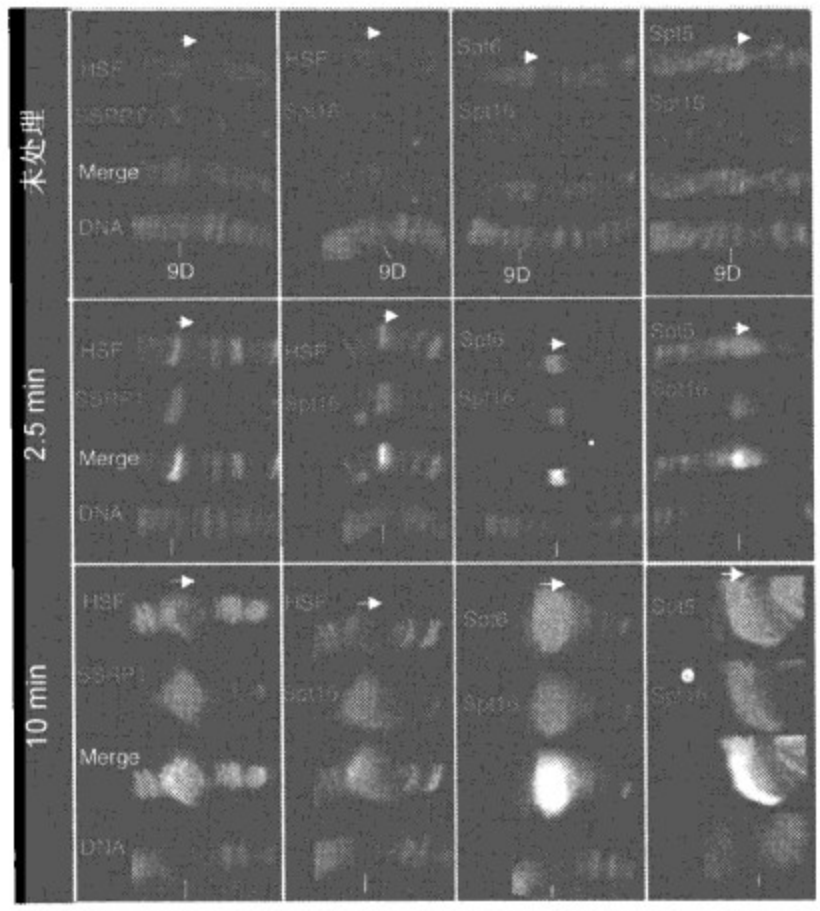


图 13.40　FACT 被迅速地召集到被转录的基因上，定位于结合了启动子的激活因子的下游。Reinberg 和 Lis 及同事用荧光抗体进行果蝇染色体染色，细胞分别是未处理，热激 2.5min、10min，如图的左边所示。所用抗体在每个染色的染色体旁边标出，与荧光抗体的颜色相同。对 HSF 和 Spt16 的专一性抗体发绿色荧光，对 SSRP1 和 Spt16 的专性抗体发红色荧光。将这两种图像合并，检测其重叠情况。有重叠时显黄色，重叠不完全时，在黄色的右边（下游）会出现一些红色荧光，这种情况在热激 10min 后的 HSF 和 SSRP1 叠加时尤其明显（左下图）。染色体还用 Hoechst 染料染色，使染色体呈现紫色（每个图的下面）。（*Source*: Reprinted with permission from *Science*, Vol. 301. Abbie Saunders, Janis Werner, Erik D. Andrulis, Takahiro Nakayama, Susumu Hirose, Danny Reinberg, and John T. Lis. "Tracking FACT and the RNA Polymerase II Elongation Complex Through Chromatin in Vivo," Fig. 2, p. 1095. Copyright 2003. AAAS.）

相比之下，当另一个推测的染色质重建物 Spt6 用绿色荧光标签染色时，它的位置和 FACT 完全一致（见第 3 个图，热激 5min 或 10min 的黄色条带）。这一结果表明，Spt6 和 FACT 都与 RNA 聚合酶 II 一起移动，处于帮助重建染色质促进转录的位置。

小结　FACT 是转录延伸促进物，由两个亚基 Spt16 和 SSRP1 组成。Spt16 结合 H2A-H2B 二聚体，而 SSRP1 结合 H3-H4 四聚体。FACT 通过推动核小体上至少一个 H2A-H2B 二聚体的丢失而促进转录通过核小体。FACT 也具有组蛋白分子伴侣作用，可将 H2A-H2B 二聚体再加回失去该二聚体的核小体。FACT 的 Spt16 亚基具有酸性 C 端，对核小体的这两种重构活性都很重要。

总结

真核生物 DNA 与叫做组蛋白的碱性蛋白质分子结合，形成核小体结构。核小体包含 4 对核心组蛋白（H2A、H2B、H3、H4），它们构成楔形盘状结构，周围包裹着约 147bp 的 DNA 片段。组蛋白 H1 比核心组蛋白更容易从染色质上去除，因而不是核心核小体的组成部分。

染色质折叠的二级结构是 30nm 纤丝。关于这种纤丝结构的线索来自四聚核小体的晶体结构，是由双摞核小体组成，两摞之间有 Z 字形的连接体 DNA。这种排列与螺线管模型以及多数其他关于 30nm 纤丝的模型都不同，但与交叉连接体双起始螺旋一致，即两摞核小体在左手双螺旋中可以相互缠绕。

染色质压缩的三级结构涉及真核细胞染色质的放射型环状结构的形成。30nm 纤丝进行某种形式的向回折叠或螺旋而形成 50nm 纤丝。然后，50nm 纤丝形成 35～85kb 长的环状结构，锚定在染色体的中央基质上。

非洲爪蟾 5S rRNA 基因的两个家族分别是卵母细胞基因和体细胞基因。前者仅在卵母细胞中表达，后者在卵母细胞和体细胞中均可表达。造成这种差异的主要原因在于体细胞基因与转录因子形成更为稳定的复合物。转录因子通过阻止核小体与体细胞基因内部调控区形成稳定复合物而使基因保持活性。形成稳定的抑制复合物需要组蛋白 H1 的参与，一旦此类复合物形成，转录因子就被排除在外，从而抑制基因活性。

核心组蛋白（H2A、H2B、H3、H4）能在裸 DNA 上组装成核小体核心颗粒。当重建染色质达到每 200bp DNA 一个核小体核心颗粒的密度时，75% 的 II 类基因的转录将被抑制（相对于裸 DNA）。剩余 25%的转录活性是由启动子区未被核小体核心颗

粒覆盖所引起的。组蛋白 H1 能在核心核小体的基础上更进一步的抑制模板活性，这种抑制作用能被转录因子阻止。有一些转录因子，如 SD1 和 GAL4，既是阻遏物（阻止由组蛋白 H1 引起的抑制作用），又是转录激活因子。其他的仅仅是阻遏物，如 GA-GA 因子，推测这些阻遏物与组蛋白 H1 在 DNA 模板上的同一位点存在竞争关系。

活跃基因具有 DNase 超敏感调控区。这种超敏感性至少部分是因为核小体的缺乏产生的。

在细胞质和细胞核中均可发生组蛋白乙酰化。在细胞质的乙酰化由 HAT B 执行，并为核小体的装配准备组蛋白，这些乙酰基团随后在细胞核中被除去。在核内的乙酰化由 HAT A 催化，并与转录激活有关。许多辅激活因子有 HAT A 活性，能够松弛核小体与基因调控区的结合。核心组蛋白尾部的乙酰化也吸引溴基区蛋白如 $TAF_{II}250$，该蛋白质对转录至关重要。

转录抑制剂如无配体核受体和 Mad-Max 结合到 DNA 位点上，并与 NCoR/SMRT、SIN3 等辅阻遏物相互作用，然后与组蛋白脱酰酶（如 $HDAC_1$ 和 $HDAC_2$）结合。这种三元蛋白复合物的装配使组蛋白脱酰酶更接近邻近的核小体。核心组蛋白的去乙酰化使组蛋白的碱性尾紧密结合到附近核小体的 DNA 和组蛋白上，稳定核小体，抑制转录。

许多真核基因的激活需要染色质重建。几种不同的蛋白质复合物执行这一重建任务，它们都有 ATPase 活性，可以控制 ATP 水解，为染色质重建提供能量。依据 ATPase 组分的不同可区分重建复合物。其中两个研究最清楚的复合物是 SWI/SNF 和 ISWI。哺乳动物的 SWI/SNF 复合物以 BRG1 为其 ATPase，还有 9～12 个 BRG1-相关因子（BAF）。其中一个高度保守的 BAF 叫做 BAF155 或 BAF170，它具有负责组蛋白结合的 SANT 结构域，可帮助 SWI/SNF 结合到核小体上。重建复合物的 ISWI 家族具有 SANT 结构域和另一个叫做 SLIDE 的结构域，SLIDE 结构域可能参与 DNA 结合。

对染色质重建机制的细节还不清楚，但确实涉及核小体的移动，伴随 DNA 与核心组蛋白间连接的松弛。与核小体中非催化 DNA 的暴露，或核小体沿一段 DNA 简单滑行不同，被催化核小体的重建涉及与核心组蛋白对应的核小体 DNA 不同构象的形成。

ChIP 分析可以揭示在激活期间与基因结合的蛋白质因子的顺序。酵母 *HO* 基因活化时，首先结合的是 Swi5，接着是 SWI/SNF 和 SAGA，后者含有 HAT Gcn5p，然后是通用转录因子和其他蛋白质。因此，染色质重建位于该基因激活的第一步，但是不同基因的蛋白质结合顺序也不尽相同。

在特定核小体中，核心组蛋白的修饰方式组成了组蛋白密码。组蛋白密码能决定核小体里发生什么事件。IFN-β 增强体的激活因子能召集一种组蛋白乙酰转移酶，该酶使启动子处的核小体中组蛋白 H3、H4 的一些赖氨酸乙酰化。一种蛋白激酶也能使同一核小体的组蛋白 H3 的一个丝氨酸磷酸化，由此允许组蛋白 H3 的另一个赖氨酸乙酰化，完成组蛋白编码。组蛋白 H4 的一个赖氨酸被乙酰化后，接着召集 SWI/SNF 复合物，重构核小体。然后 TFIID 与组蛋白 H3 上的两个乙酰化赖氨酸结合。TFIID 的结合使 DNA 弯曲，将重构的核小体移开，为起始转录开通道路。

常染色质相对伸展，具有潜在活性。而异染色质处于凝缩状态，没有遗传学活性。异染色质能使 3kb 以内的基因发生沉默。酵母染色体末端（端粒）异染色质的形成依赖于蛋白 RAP1 对端粒 DNA 的结合，随后依次召集蛋白 SIR3、SIR4、SIR2。染色体其他位置上的异染色质形成也依赖蛋白 SIR。SIR3 和 SIR4 也直接与核小体组蛋白 H3、H4 相互作用。核小体组蛋白 H4 的 Lys16 乙酰化阻止其与 SIR3 的相互作用，因而抑制异染色质形成。这是组蛋白乙酰化促进基因活性的另一种方式。

组蛋白 H3 N 端尾的 Lys9 甲基化吸引蛋白质 HP1，随后 HP1 召集组蛋白甲基转移酶（可能使邻近核小体上的 Lys9 甲基化），从而扩展被抑制及异染色质化的状态。核心组蛋白中其他赖氨酸和精氨酸侧链的甲基化有抑制或激活双重效应，这些甲基化与其他组蛋白修饰方式，包括乙酰化、磷酸化、泛素化等同时发生在特定核小体上。一般情况下，每种特定的修饰组合能将不同的转录抑制或激活信息传递给细胞。特定的组蛋白修饰对其他或邻近的修饰方式也会产生影响。

FACT 是转录延伸促进物，由两个亚基 Spt16 和 SSRP1 组成。Spt16 结合 H2A-H2B 二聚体，而 SSRP1 结合 H3-H4 四聚体。FACT 通过使核小体上至少一个 H2A-H2B 二聚体的丢失而促进转录穿越核小体。FACT 还具有组蛋白分子伴侣作用，可将 H2A-H2B 二聚体再加回失去该二聚体的核小体。FACT 的 Spt16 亚基具有酸性 C 端，该末端对核小体的这两种重构活性都很重要。

复习题

1. 依照下列要求图示核小体：（a）无 DNA 分子的核小体组蛋白，显示所有组蛋白的大概位置；（b）在另一图中显示组蛋白上 DNA 缠绕的路径。

2. 用电镜证据说明 DNA 分子在核小体中被压

缩为原先的 1/7～1/6。

3. 用电镜证据说明在高离子强度下，染色质浓缩形成 30nm 纤丝。

4. 绘制由 X 射线晶体学揭示的四聚核小体的结构。

5. 四聚核小体结构提示 30nm 纤丝具有怎样的结构?

6. 绘图阐明 30nm 纤丝之后的下一级染色质折叠结构。引用生化和显微镜证据支持该模型。

7. 绘图阐明转录因子和组蛋白 H1 在重建染色质的 5S rRNA 基因转录中的竞争效应。

8. 叙述实验并给出结果，说明组蛋白 H1 和激活因子 GAL4-VP16 在重建染色质的腺病毒 E4 基因转录的中的竞争效应。

9. 给出两个模型说明转录激活因子的抗阻抑作用。其中一个的基因调控区无核小体，另一个则与核小体结合。

10. 叙述实验并给出结果，显示在活跃的 SV40 染色质中，无核小体区位于病毒晚期基因的调控区。

11. 叙述实验并给出结果，显示在 SV40 染色质中的 DNase 超敏感区位于病毒晚期基因的调控区。

12. 图示并描述一种检测 DNase 超敏感区的常规技术。

13. 描述并给出活性胶测定实验结果，显示 HAT 活性的存在。

14. 给出模型，说明在转录抑制中涉及辅助抑制物和组蛋白去乙酰基酶。

15. 描述并给出抗原表位标签实验结果，并显示下列 3 种蛋白质的互作：抑制物 Mad1、辅助抑制物 SIN3A、组蛋白去乙酰化酶 HDAC2。

16. 给出模型，说明同一种蛋白质在配体存在与否条件下对应的激活和抑制作用。

17. 给出模型，说明未催化核小体 DNA 暴露和催化核小体的重建。给出这一催化模型的证据。

18. 描述如何利用染色质免疫沉淀法检测细胞周期不同时点上与特定基因结合的蛋白质。

19. 叙述实验并给出结果，显示使用染色质免疫沉淀技术发现 IFN-β 启动子结合的核小体中，核心组蛋白特殊位点的乙酰化和磷酸化时间进程。

20. 叙述实验并给出结果，检测 IFN-β 启动子结合的野生型或突变型组蛋白对 SWI/SNF 和 TFIID 的召集。

21. 给出模型，描述在 IFN-β 启动子上组蛋白密码的建立和解读。

22. 给出模型，解释为什么组蛋白 H4 的 Lys16 对基因沉默很关键? 有什么证据支持这一假说?

23. 给出模型，描述组蛋白甲基化促进染色质的抑制作用。

24. 给出模型，显示组蛋白 H3 N 端尾部的 Lys9、Lys14 及 Ser10 的修饰之间的相互作用。同时显示激活和抑制相互作用。

25. 给出证据，证明 FACT 引起核小体上 H2A-H2B 二聚体的丢失，并且这种活性依赖于 FACT 的 Spt16 亚基的 C 端。

分析题

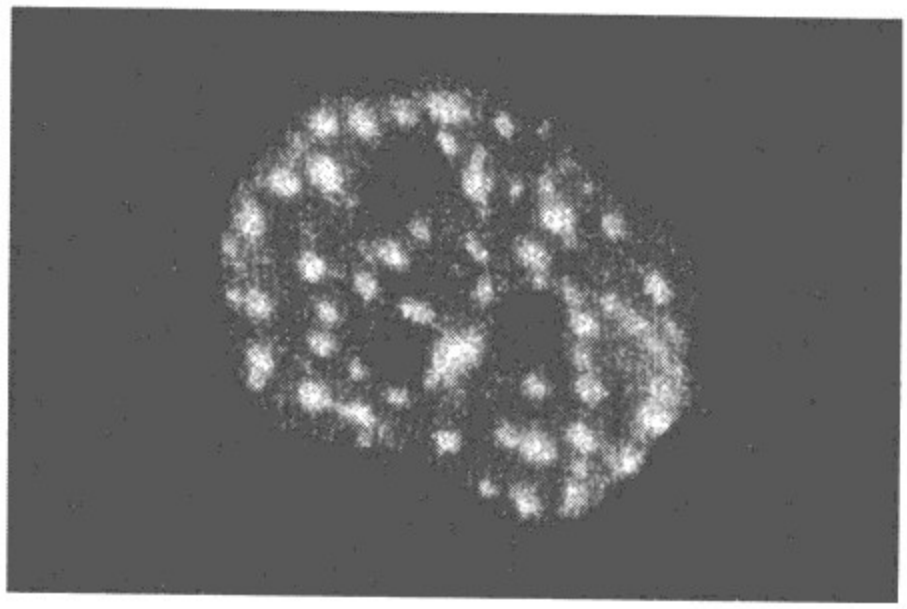
1. 如果 J6 细胞的球蛋白基因位点具有与 HEL 细胞一样的 DNase 超敏感位点，按图 13.22d 所示，约多大的片段能被检测到? 哪些超敏感位点不能被检测到?

2. 解释为什么利用微球菌核酸酶对真核染色质短时间消化时，会产生 200bp 大小的 DNA 片段，而较长时间消化却产生 147bp 的片段。

3. 核心组蛋白的氨基酸序列在动植物中高度保守，提出假说解释这一现象。

翻译　马　纪　兰海燕　校译　张富春

第14章　mRNA加工I：剪接



细胞核内人类前体mRNA剪接因子的电子显微镜图像（×14 500）。© *Dr. Darid Spector/Peter Arnold, Inc.*

细菌基因的表达可简要总结如下：首先，RNA聚合酶转录操纵子中的一个或一套基因；然后，甚至转录仍在进行时，核糖体就结合到mRNA分子上进行蛋白质的翻译。我们在第6～8章已经学习了该过程中的转录部分，并且看上去十分复杂。然而，真核生物的转录过程就更为复杂了。在第13章我们已经讨论过染色质的复杂结构，但其复杂性不仅仅局限于此。

真核生物的转录和翻译在细胞中所发生的部位不同。转录发生在细胞核内，而翻译在细胞质中进行。这意味着转录和翻译不可能像细菌那样同时发生。转录完成后，转录物由细胞核转运到细胞质，翻译才能进行。转录与翻译在时间上存在一定的间隔，因此把翻译称为转录后阶段（posttranscriptional phase）。

在本章我们会了解到，与典型的细菌基因相比，大多数真核生物基因的编码区被非编码DNA所隔断。RNA聚合酶不能区分基因的编码区和非编码区，它会转录基因的全部序列。因此，细胞必须通过剪接过程，将初级转录物中非编码RNA去除。真核生物还会在其mRNA的5′端和3′端添加特殊结构，即5′端的帽子结构、3′端由AMP串联形成的poly（A）结构。这些mRNA的加工事件均在细胞核内完成，且发生于转录结束前。因此，将这些加工事件称之为共转录事件（cotranscriptional event）更为准确，而不应称之为转录后事件（posttranscriptional event）。为避免混淆，我们将剪接、5′端加帽、3′端加尾通称为mRNA加工事件。在详细学习了剪接（本章）、加帽和多聚腺苷化（第15章）后，在第15章结束时，再进行相关论题的探讨。

14.1　断裂基因

如果我们用一个句子来表达人类β-球蛋白基因的序列，可采用以下表达方式。

这是*bhgty*人类β-球蛋白*qwtzptlrbn*基因。

很显然，基因内由斜体英文字母构成的2个区域（*bhgty*和*qwtzptlrbn*）没有任何意义。这两个序列与其两侧编码人类β-球蛋白的基因序列完全不相干。有时将这些序列称之为**插入序列**（intervening sequence）或**IVS**，但通常遵循Walter Gilbert的命名，称为**内含子**（intron）。同理，基因内有意义的那些区域被称为编码区或表达区，但Gilbert将其称为**外显子**（exon），且这一称谓被普遍采用。但有些基因特别是低等真核生物的基因根本就没有内含子区，而有些基因却含有大量的内含子。例如，编码人类巨肌蛋白的*titin*基因是迄今发现的含有内含子最多（362个内含子）的基因。

有关断裂基因的证据

1977年，Phillip Sharp及同事首次在腺病毒的主要晚期基因座上发现了内含子。腺病毒的主要晚期基因座上的基因可在侵染的后期进行转录，编码结构蛋白如病毒衣壳蛋白组分之一的六聚体蛋白（hexon）。有证据表明腺病毒的主要晚期基因座上的基因是断裂基因，利用**R环**（R-looping）实验技术可以很容易地证明这一点。

在R环实验中，首先使RNA分子与其模板DNA单链分子杂交，也就是说，先将DNA模板双链分开，使模板链与RNA链进行杂交形成RNA-DNA双链杂合分子。事实上，在实验条件下，这种RNA-DNA双链杂合分子比双链DNA分子更为稳定。然后用电子显微镜观察杂交分子的特征。开展该项实验的基本途径有两条：①使双链DNA解旋的长度仅够与RNA杂交即可；②在杂交之前，使DNA双链完全解离。Sharp及同事采用的是后一种方法，即先使腺病毒的DNA双链分子解旋分离成单链，然后与编码腺病毒六聚体蛋白的成熟mRNA杂交，图14.1给出了实验结果（注意不要因exon和hexon写法相似而将这两个术语混淆，它们之间毫不相干）。

如果六聚体蛋白基因没有内含子，则mRNA与其DNA模板因互补配对而整齐排列，形成平滑的线形杂交分子。但是，如果六聚体蛋白基因含有内含子，其情况又会如何呢？很显然，成熟的mRNA分子没有内含子，否则内含子编码的无义信息就会出现在蛋白质产物中。也就是说，DNA上的内含子序列在成熟mRNA分子上缺失。这就意味着，当六聚体蛋白基因的DNA分子与其mRNA杂交时，不可能形成平滑的线形杂交分子。取而代之的是，DNA的内含子序列因mRNA上无与之互补序列而形成环，在图14.1所示的实验中确实有环形成。该环由DNA序列形成，但由于是与RNA杂交而形成的，因此称为R环。

电子显微镜图像显示，RNA-DNA杂交分子被3个单链DNA环（在图14.1中标注为A、B和C）所间隔，这些DNA环代表*hexon*基因的内含子。每环的前面都有一短的杂交区，而最后一个环的后面是长的杂交区。由此可见，该基因包含4个外显子：位于基因初始区域附近的3个较短的外显子，其后为一较长的外显子。3个短外显子转录为*hexon*基因mRNA 5′端的前导区，位于编码区之前。而长外显子是*hexon*基因的编码区。事实上，腺病毒的主要晚期基因含有不同的编码区，但所有都具有由这3个短外显子编码的前导区。

发现病毒基因具有内含子这一惊人事件后，我们想知道，这是否只是病毒的一种不同寻常现象而已，与真核生物细胞的生命过程没有关联。因此，有必要确定真核细胞内的基因是否也具有内含子。Chambon及同事以鸡卵清蛋白基因为研究对象，进行R环实验分析以确定该基因是否具有内含子。研究者观察到6个因不能与mRNA杂交而形成大小不同的DNA环，说明鸡卵清蛋白基因含有7个外显子和6个内含子。有趣的是，多数内含子都比外显子长，内含子在数量上占优势是高等真核生物基因的显著特征，低等真核生物（如酵母）基因内的内含子序列一般较短，且数量较少。

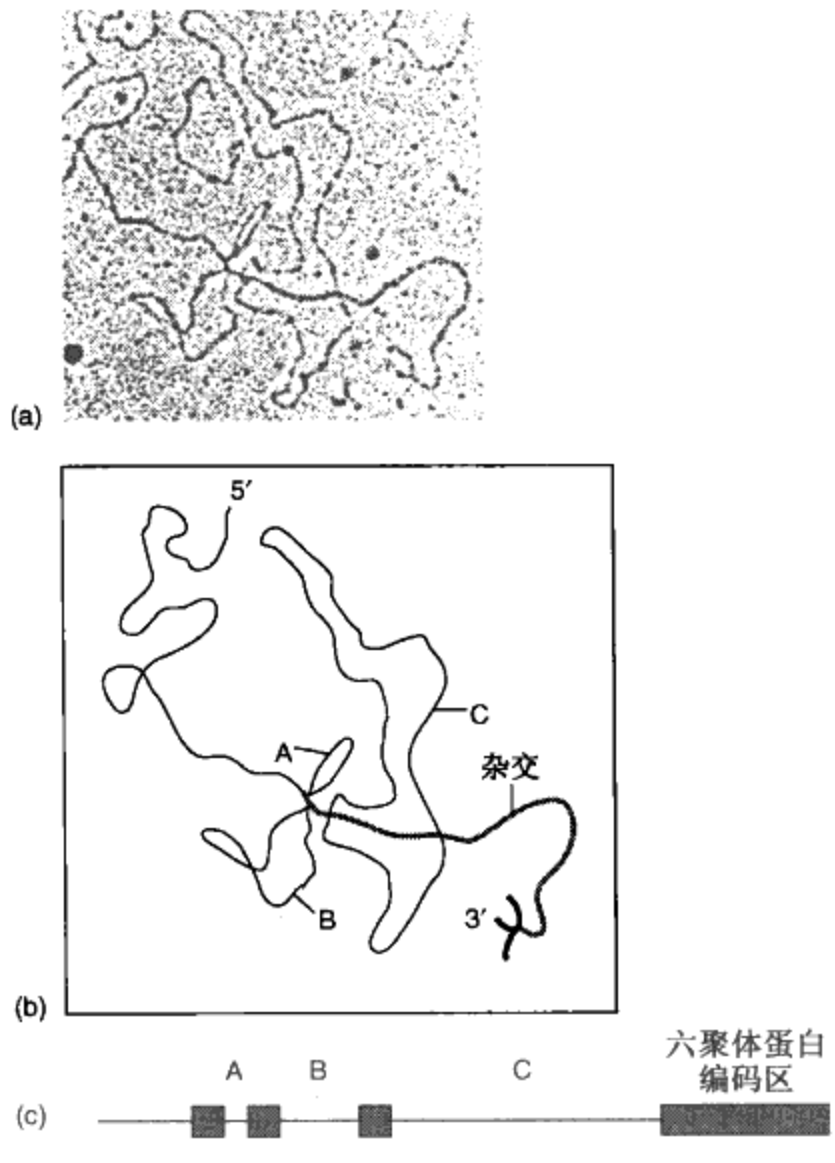


图14.1　R环实验揭示腺病毒基因存在内含子。

（a）包含晚期六聚体蛋白基因5′端序列的腺病毒DNA克隆片段与成熟mRNA杂交的电子显微图像。环代表DNA分子中不能与mRNA杂交的内含子序列。（b）电子显微图像的注释简图。显示3个内含子环（A、B、C）、杂交区域（红色粗线）和基因上游未参与杂交的DNA区域（左上）。右下方的叉形结构由mRNA未杂交的3′端序列形成，因为参与杂交的DNA分子不包含*hexon*基因的3′端序列。因此mRNA在3′端形成分子内双链结构。（c）*hexon*基因的线性排列图。包含3个较短的被2个内含子（A、B）所隔离的前导外显子，一个较长的内含子（C）将前导区与编码区隔开。所有的外显子用红色方块表示。（*Source*：（a）Berget，M.，Moore，and Sharp，Spliced Segments at the 3′ terminus of adenovirus 2 Late mRNA *proleedings of the National Academy of Sciences USA* 74：3173，1977.）

到目前为止，我们只讨论了mRNA基因的内含子，但有些tRNA基因甚至rRNA基因也有内含子，

但与 mRNA 基因的内含子相比，存在少许差异。例如，tRNA 基因的内含子相对较小，长度范围为 4～50bp，有些 tRNA 基因根本就没有内含子，而有内含子的 tRNA 基因只有一个内含子，与编码反密码子的 DNA 碱基相邻。线粒体和叶绿体内的基因也有内含子。事实上，正如我们将要了解到的，这些内含子是最令人感兴趣的 DNA 序列。

小结 在高等真核生物的细胞内，大多数编码 mRNA、tRNA 以及少数编码 rRNA 的基因其编码区被一些称为内含子的不相干序列所分割。内含子两侧的序列为外显子，最终会在出现在 RNA 产物中。编码 mRNA 的基因含有 0～362 个数目不等的内含子，而 tRNA 基因要么没有内含子，要么只有 1 个内含子。

RNA 剪接

现在我们来考虑因内含子存在而引发的问题。存在于基因中的内含子并没有出现在成熟 mRNA 分子中，内含子的信息如何缺失于基因的成熟 RNA 产物？主要有两种可能：一种是内含子根本就没被转录，RNA 聚合酶在执行转录功能时，以某种方式从一个外显子跳到另一个外显子，忽视内含子的存在；另一种是内含子也被转录，产生初级转录物，过大的基因产物通过加工过程切除内含子。看起来是一种浪费的过程，但第二种可能是正确的。将内含子切除、外显子拼接形成成熟 RNA 的过程，称为 **RNA 剪接**（RNA splicing）。图 14.2 概述了剪接的过程，在后面我们会发现该图太过简略。

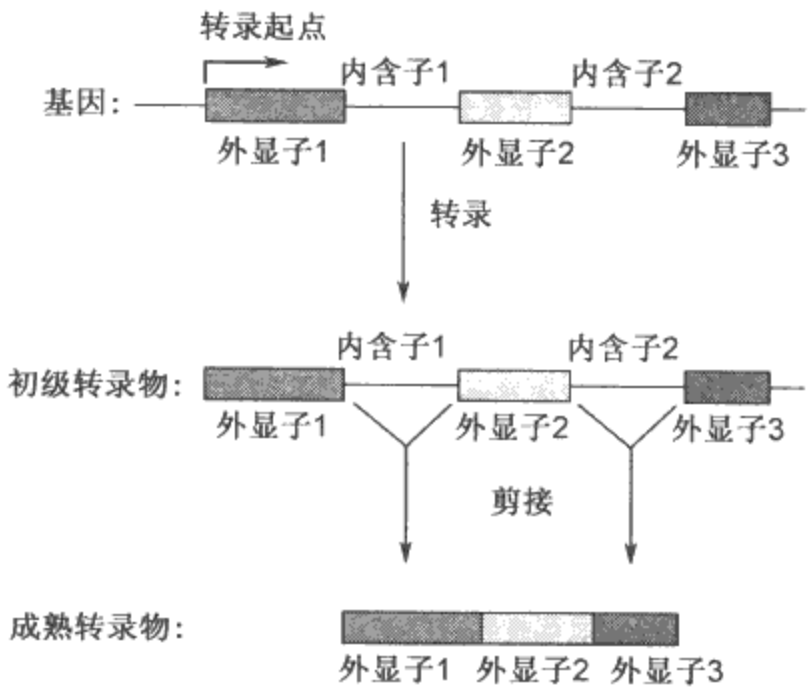


图 14.2 RNA 剪接的要点。基因中的内含子与外显子（带颜色的方块）共同被转录，形成初级转录物，然后切除内含子、拼接外显子。

我们如何确知剪接事件的发生？其实，早在内含子发现之时，就有间接证据表明剪接事件的存在。核内不均一 RNA（hnRNA）（第 10 章）的发现为 RNA 剪接事件提供了证据，这类定位于细胞核内的较大 RNA 分子被认为是 mRNA 前体。hnRNA 的大小（比 mRNA 大）和定位（细胞核）与未经剪接的 mRNA 前体完全一致，而且，hnRNA 很快会转化为比较稳定的小 RNA，这暗示 hnRNA 仅仅是形成更稳定 RNA 的中间体。然而，还没有直接的证据表明 hnRNA 能够被剪接形成 mRNA。

小鼠 β-球蛋白 mRNA 及其前体为 hnRNA 经过剪接形成 mRNA 提供了直接证据。作为 hnRNA 的一员，小鼠 β-球蛋白基因 mRNA 前体只存在于细胞核内，其长度（1500bp）是成熟 mRNA 长度（750bp）的 2 倍，且很快就转化为较小的 RNA 分子。此外，小鼠红细胞可合成数量众多的球蛋白（占细胞蛋白质总量的 90%），因此该细胞含有大量的 α-球蛋白及 β-球蛋白 mRNA，甚至 mRNA 前体的含量也十分丰富，所以，进行 mRNA 的纯化相对比较容易，使开展 RNA 剪接分析成为可能。而且，β-球蛋白前体的大小与包含有内含子和外显子的初级转录物的大小完全一致。Charles Weissmann 和 Philip Leder 及同事通过 R 环验证了前体含有内含子的假说。

进行 R 环实验时，分别用球蛋白的成熟 mRNA 或 mRNA 前体与克隆的球蛋白基因进行杂交，然后观察 R 环的形成情况（图 14.3）。我们能够预知成熟 mRNA 与基因的杂交结果，因为成熟 RNA 分子无内含子序列，因此，基因内的内含子序列就会形成 R 环而突显出来。此外，如果前体 RNA 含有全部内含子序列，在与基因杂交时不会有 R 环形成，实验结果也确实如此。由于在 R 环实验中使用的是双链而不是单链 DNA，杂交时 RNA 链置换了其中的一条链而与另一条 DNA 链杂交。因此，对图 14.3 中的结构识别起来可能较为困难。前体 RNA 与 DNA 杂交后形成平滑的杂交分子，无 R 环的形成；而成熟 RNA 与 DNA 杂交后会产生 R 环，被明显的双链 DNA 环所间断，该 DNA 环对应于基因内的一个大内含子。在本实验中观察不到小的内含子。值得注意的是，“内含子”一词也可用于表示 DNA 和 RNA 分子内的插入序列。

小结 真核生物信使 RNA 的合成是分阶段进行的。第一阶段是合成初级转录物 mRNA 前体，即含有基因内含子的拷贝序列，这些 mRNA 前体是 hnRNA 群体中的一部分。第二阶段是 mRNA 的成熟，切除内含子的剪接过程是 mRNA 成熟事件的重要组成部分，通过剪接形成成熟的 mRNA。

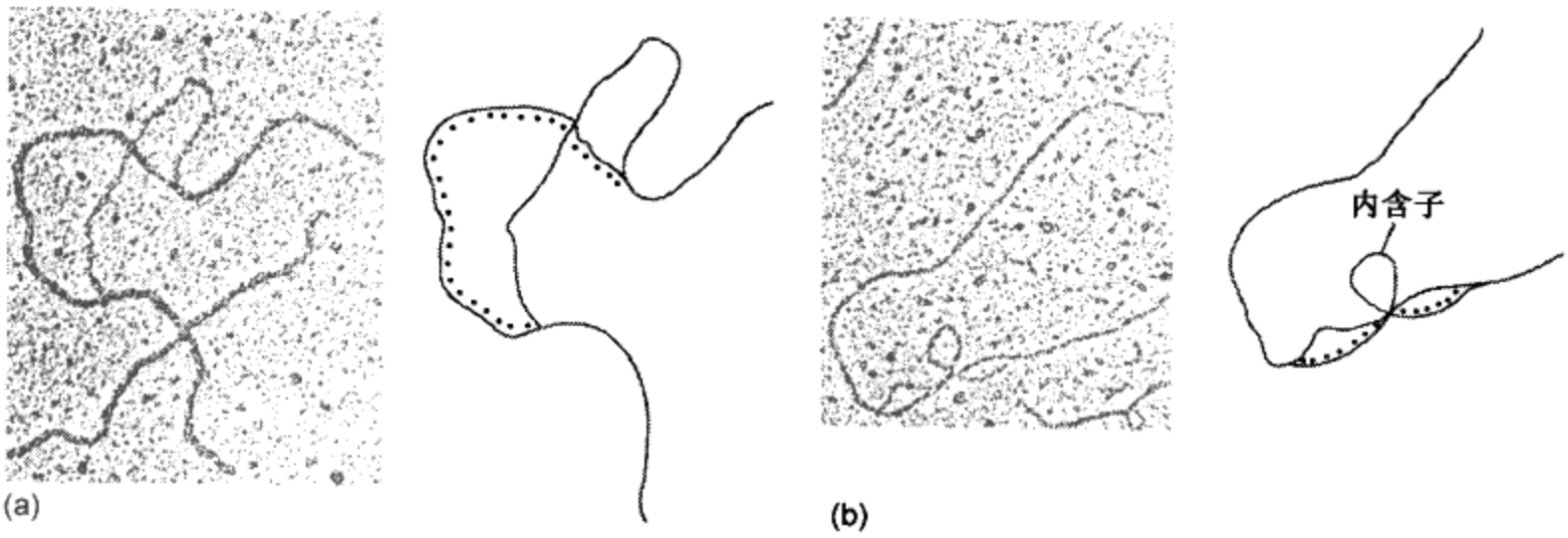


图 14.3　内含子可被转录。(a) 在 R 环实验中，小鼠球蛋白 mRNA 前体与 β-球蛋白的克隆基因杂交形成平滑的杂交分子，说明内含子存在于 mRNA 前体中。(b) 用成熟的球蛋白 mRNA 进行相同的 R 环实验时，杂交分子出现了由基因内含子形成的环，这表明成熟 mRNA 没有内含子。本实验未能观察到小内含子。在注释图中，黑点线表示 RNA，红线表示 DNA。(*Source*: Ttilghman, S., P. J. Curtis, D. C. Tiemeier, P. Leder, and C. Weissmann, The intervening sequence of a mouse β-globin gene is transcribed within the 15S β-globin mRNA precursor. *Proceedings of the National Academy of Sciences USA* 75: 1312, 1979.)

剪接信号

mRNA 前体的准确剪接是十分重要的。如果从 mRNA 前体中切除的 RNA 片段太小，那么成熟的 RNA 分子就会被无义区域所间断；而如果从mRNA 前体中切除的 RNA 片段太大，那么成熟的 RNA 分子就会丢失一些重要的序列。

既然准确剪接如此重要，那么，在 mRNA 前体中必定存在着剪接信号，指导剪接体准确地对其进行“切割和连接”。那么，这些剪接信号是什么？找出剪接信号的方法之一就是分析众多不同基因内含子边界的碱基序列，找出所有基因内含子边界处的共有序列。从理论上讲，这些共有序列可能是mRNA前体剪接信号的一部分。Chanbon 等首次获得了最为引人注目的发现是，几乎所有细胞核 mRNA 前体的内含子都有相同的开始和结束序列：

外显子/GU-内含子-AG/外显子

也就是说，转录物内含子 5′端最初的两个碱基为 GU，3′端最末尾的两个碱基为 AG，这种保守性绝非偶然，GU-AG 基序一定是使剪接体找到准确剪切位点的部分剪接信号。然而，GU-AG 基序出现的概率如此频繁，以至于一个典型的内含子中会包含多个 GU 和 AG 序列，为什么这些序列不能作为剪切位点？答案在于，GU-AG 基序并不是剪接信号的全部，真正的剪接信号比 GU-AG 基序更为复杂。目前已对众多基因的外显子-内含子间的边界序列进行了分析，在哺乳动物中，外显子与内含子的边界保守序列为：

5′-AG/GUAAGU-内含子-YNCUR A̲C-Y_nNAG/G-3′

其中，“/”表示外显子-内含子的边界；“Y”表示嘧啶碱基（U 或 C）；“Y_n”表示一段约由 9 个嘧啶核苷酸组成的序列；“R”表示任意一种嘌呤碱基（A 或 G）；“A̲”为参与形成分支剪接中间体的特殊腺苷酸，相关内容将在本章后面介绍；“N”表示任意碱基。对酵母 mRNA 前体的外显子-内含子边界的保守序列研究发现，与哺乳动物的外显子-内含子间的边界序列相比，具有细微的差别：

5′-/GUAUGU-内含子-UACUA A̲C-YAG/-3′

发现保守序列后，我们还必须证明这些序列对于 mRNA 前体正确剪接确实十分重要。有几个研究小组为这些剪接保守序列的重要性提供了大量证据。这些研究小组所开展的实验基本分为两类：第一类是在克隆基因的剪接处对保守序列进行突变，然后分析剪接是否仍能正常发生；第二类是从那些被怀疑是剪接出现问题的患者中收集缺陷突变基因，分析这些基因在剪接位点处的突变情况。两类研究都得出了相同的结论：保守序列遭到破坏后通常会抑制正常的剪接。

尽管外显子边界的剪接信号是十分必要的，但这些边界序列还不足以界定一个外显子。在下一节我们将介绍邻近内含子末端的序列（称为分支点序列）也是下一个外显子被识别的必需序列，甚至位于分支点序列连同剪接信号也不足以界定一个外显子，这是因为高等真核生物中有许多内含子都很大，其长度可达 100kb，其内包含有许多外显子大小的序列，这些序列的边界也具有类似于正常剪接信号的序列特征，包括分支点序列。然而，在极个别情况下，这些“假外显子”也可能会被剪接而出现在

成熟 mRNA 分子中。是什么特征使真正的外显子与这些假外显子区分开来？部分原因在于，真正的外显子一般包含能促进剪接的**外显子剪接增强子序列**（exonic splicing enhancer，ESE），而假外显子一般包含能抑制剪接的**外显子剪接沉默子序列**（exonic splicing silencer，ESS）。我们将在本章后面详细讨论这种现象。

小结 细胞核 mRNA 前体具有高度一致的剪接信号，内含子的最初两个碱基几乎总是 GU，而最末尾的两个碱基几乎总是 AG。在 GU 和 AG 基序两侧的 5′-剪接位点及 3′-剪接位点具有保守序列。全部的保守序列对于正确剪接十分重要，当保守序列发生突变后，会发生异常剪接。

14.2 细胞核 mRNA 前体的剪接机制

图 14.2 剪接示意图只简单地显示了前体与剪接产物，并没有涉及细胞将前体转化为成熟 mRNA 时所采用的机制。下面我们将深入探讨复杂而又意想不到的细胞核 mRNA 前体的剪接机制。

分支中间体

图 14.2 未阐明的主要细节之一就是细胞 mRNA 前体剪接时形成的中间体是有分支的，该中间体类似套索（lariat）状结构，如同放牧牛仔的索套。图 14.4 示意了通过两步转酯反应进行剪接的套索模型，第一步反应是形成套索状中间体。由内含子中部的腺嘌呤核苷酸通过 2′-羟基对第一个外显子与内含子起始碱基 G 之间的磷酸二酯键发动亲核攻击而引发的，形成套索环，同时第一个外显子被切离。第二步则完成剪接过程。以第一步释放的外显子的游离 3′-羟基作为亲核基团，攻击 3′-剪接位点处连接内含子和第二个外显子的磷酸二酯键，将两个外显子通过 5′，3′-磷酸二酯键连接起来，同时将内含子以套索结构的形式释放出来。

似乎还没有足够严密的证据使人们接受这个剪接机制。事实上，已有充分证据支持图 14.4 中所示中间体的存在，其中大部分证据是由 Sharp 及同事提供的。

首先，也是最为重要的一点是，分支中间体存在的证据是什么？1984 年，Sharp 及同事用无细胞剪接提取物于体外对只有一个内含子的 RNA 分子进行剪接时，首次发现在 RNA 剪接过程中会产生这种奇怪形状的 RNA 分子。剪接底物是以腺病毒主要晚期基因的最初几百个碱基序列为模板合成的

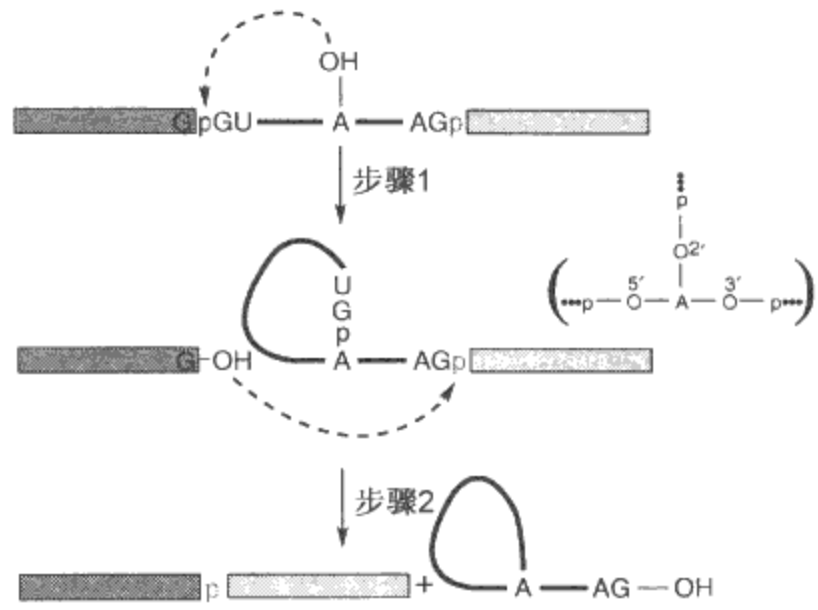


图 14.4 细胞核 mRNA 前体剪接的简化机制。步骤 1：内含子序列内的腺嘌呤核苷酸的 2′-OH 攻击连接第一个外显子（蓝色）与内含子的磷酸二酯键。这次攻击（如图上方的虚线箭头所示）破坏了外显子 1 与内含子之间的磷酸二酯键，产生游离的外显子 1 和由内含子与外显子 2 形成的套索状中间体，由内含子 5′端的 GU 与分支位点的 A 形成磷酸二酯键而产生。套索是由 RNA 前体的一部分对同一分子的另一部分发动内部攻击的结果，右侧圆括号内示意分支位点腺苷酸的 2′-OH、3′-OH 和 5′-OH 均参与了磷酸二酯键的形成。步骤 2：外显子 1 游离的 3′-OH 攻击内含子与外显子 2 之间的磷酸二酯键，从而将外显子 1 与外显子 2 连接在一起，释放出套索状内含子。注意，外显子 2 的 5′端磷酸（红色）成为两个外显子的连接分子。

放射性转录物，包含最初的两个前导外显子以及一个位于两者之间长度为 231nt 的内含子。剪接反应进行一段时间后，对 RNA 进行凝胶电泳。结果发现，除了未被剪接的前体在凝胶上产生条带外，还出现了一条电泳行为异常的条带，在 4%的聚丙烯酰胺凝胶上，其泳动速度比未被剪接前体的泳动速度快，而在 10%的聚丙烯酰胺凝胶上，其泳动速度比未被剪接前体的泳动速度慢。而这正是环形或分支 RNA，如套索 RNA 所具有的特征。

这个奇异的 RNA 分子是剪接的产物吗？它确实是剪接的产物，因为存在阻断剪接反应的抗血清以及缺少剪接反应所需的 ATP 时，该 RNA 分子就不能产生。随着剪接反应的持续进行，该 RNA 分子的量也随之积聚。在电泳时还会检测到另一个具有相同电泳行为的分子，它就是由内含子与外显子 2 形成的套索状中间体。而且，Sharp 小组开展的另一个实验（图 14.5）表明，随着剪接反应的进行，这些产物会积累的越来越多，显然，这是从前体分

子中剪切下来的套索状内含子。该实验表明还存在另一种电泳行为异常的 RNA 分子，其浓度在剪接过程的初期阶段逐渐增加，其后又表现为下降，表明这是一种剪接中间体。实质上，这是携带有套索状内含子的外显子 2。由于都呈现为套索状结构，所以该 RNA 分子和内含子均表现异常的电泳行为。

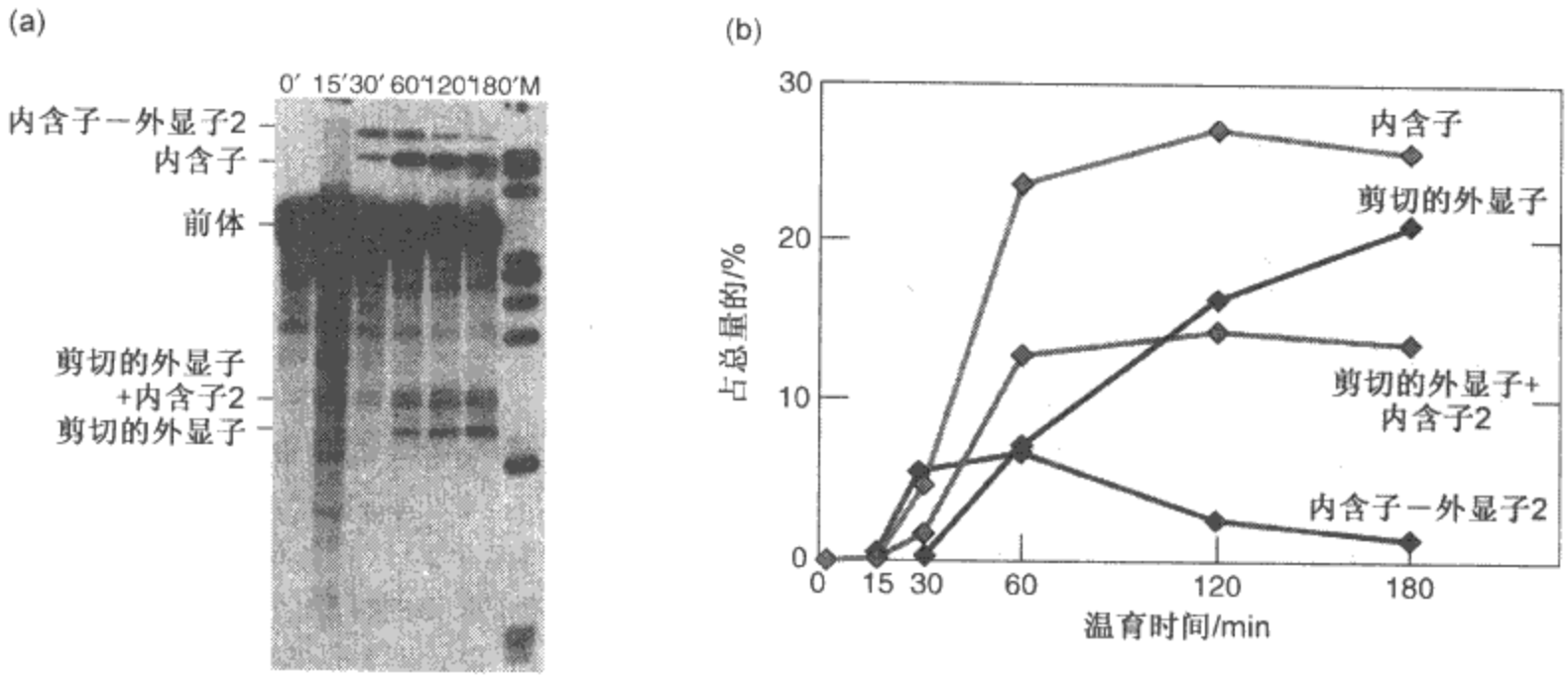


图 14.5　中间体产生及内含子释放的时间进程。（a）电泳。Sharp 及同事进行 RNA 的体外剪接反应实验，在不同的时间内将剪接反应生成的产物进行 10%的聚丙烯酰胺凝胶电泳，鉴定的产物标注在图的左侧。最顶部的条带为内含子－外显子 2 中间分子，第二条带为内含子，这两类 RNA 均为套索状，其泳动速度异常低。第三条带为 RNA 前体分子，最下面的两条带为两种剪接的外显子：上面的条带表示外显子连接有部分内含子 2 的片段，下面的条带则表明外显子无额外的内含子片段附着。（b）曲线图示。Sharp 及同事根据图（a）中每个条带的亮度，总结出了每类 RNA 分子积聚的量与时间之间的函数关系。（*Source*：Grabowski P.，R. A. Padgett，and P. A. Sharp，Messenger RNA splicing in vitro：An excised intervening sequence and a potential intermediate. *Cell* 37（June 1984）f. 4，p. 419. Reprinted by permission of Elsevier Science.）

根据图 14.4 的两步剪接模型可做出如下已被 Sharp 及同事所证实了的推论。

1. 切离的内含子具有 3′-羟基基团。例如，步骤 2 的开始阶段所示，这是外显子 1 进攻磷酸二酯键时所要求的，攻击的结果是将内含子 3′端的磷酸基团切除，而只留下羟基基团。

2. 剪接产物中，2 个外显子之间的磷原子来自 3′-（下游）剪接位点。

3. 中间体（外显子 2＋内含子）和被剪下的内含子都包含有一个分支点核苷酸，其 2′-OH、3′-OH、5′-OH 均参与了与其他核苷酸的连接。

4. 该分支涉及内含子的 5′端与内含子内部位点的结合。

下面，让我们来了解一下有关分支核苷酸存在的证据。中间体（外显子 2＋内含子）和被剪下的内含子均含有一个分支核苷酸，其 2′-OH、3′-OH、5′-OH 均参与了与其他核苷酸的连接。Sharp 及同事用 RNase T2 或 RNase P1 切割剪接中间体，两种酶均能对 RNA 分子中的核苷酸进行逐个切割，但同 RNase T1 一样，RNase T2 的酶切产物为核苷-3′-磷酸（图 14.6），而 RNase P1 的酶切产物为核苷-5′-磷酸。两种酶切反应除均能产生正常的单磷酸

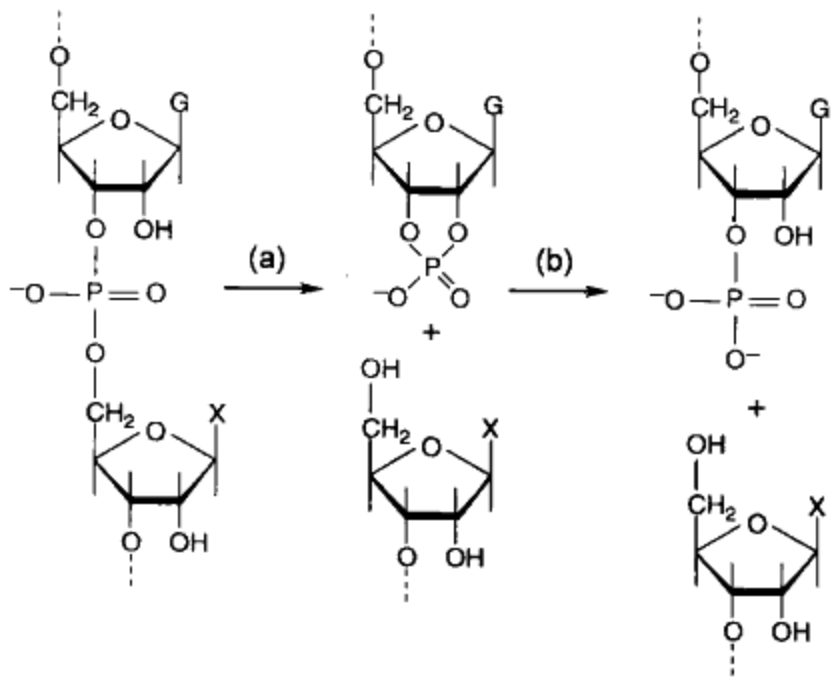


图 14.6　RNA 酶 T1 和 T2 的作用原理。这些 RNase 按下列方式切割 RNA。（a）RNase 将鸟苷酸与下游紧邻核苷酸之间的 3′，5′-磷酸二酯键打开，形成 2′，3′-环核苷酸中间体。（b）环状中间体被打开，产生末端为 3′-磷酸鸟苷酸的寡聚核苷酸。

核苷酸外，还产生一种新产物。利用薄层层析可区分这两种产物所带电荷的差异，T2 酶切产物的电荷为－6，P1 酶切产物的电荷为－4，普通单磷酸核苷酸所带的电荷为－2。

这些非正常产物是什么？图 14.7 显示了每种结构与其所带电荷间的对应关系，每个磷酸二酯键带 1 个负电荷，每个末端磷酸基团带 2 个负电荷。为证明这些结构的正确性，Sharp 及同事利用高碘酸盐和苯胺处理 P1 酶切产物，通过 β 消除作用（β-eliminarion）除去 2′-和 3′-核苷酸，形成的产物应为 2′，3′，5′-三磷酸核苷酸。二维薄层层析分析证明，这一产物的迁移率与 2′，3′，5′-三磷酸腺嘌呤核苷酸完全一致，也就证明了分支核苷酸为腺嘌呤核苷酸。

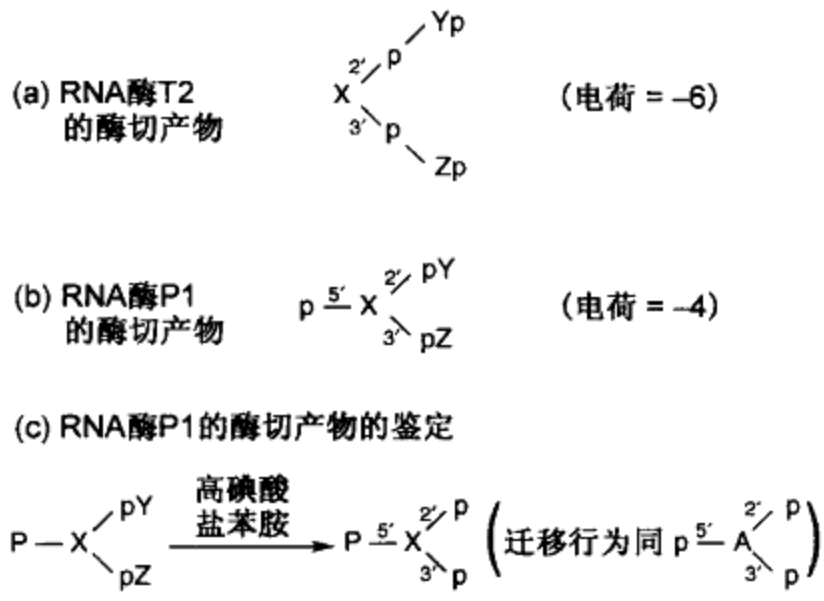


图 14.7　有关分支核苷酸的直接证据。（a）Sharp 及同事用 RNase T2 消化剪接中间体，产物带有 6 个负电荷。与图中的分支结构相一致。（b）用 RNase P1 消化的产物带有 4 个负电荷，与分支结构相一致。（c）Sharp 及同事用 RNase P1 消化的产物经高碘酸盐和苯胺处理，消除连接在分支核苷酸的 2′-和 3′-磷酸，产物为 2′，3′，5′-三磷酸腺苷酸，证明存在分支结构，且表明分支发生在腺苷酸处。

小结　一系列证据表明，细胞核 mRNA 前体的剪接要通过套索状或分支状中间体的介导。

分支点信号

参与分支的腺苷酸是否具有某些特殊之处？或许内含子内的所有腺核苷酸均具有参与分支的功能？对多个不同内含子的研究表明，在分支位点处存在保守序列，正是这个保守序列参与了分支的形成。

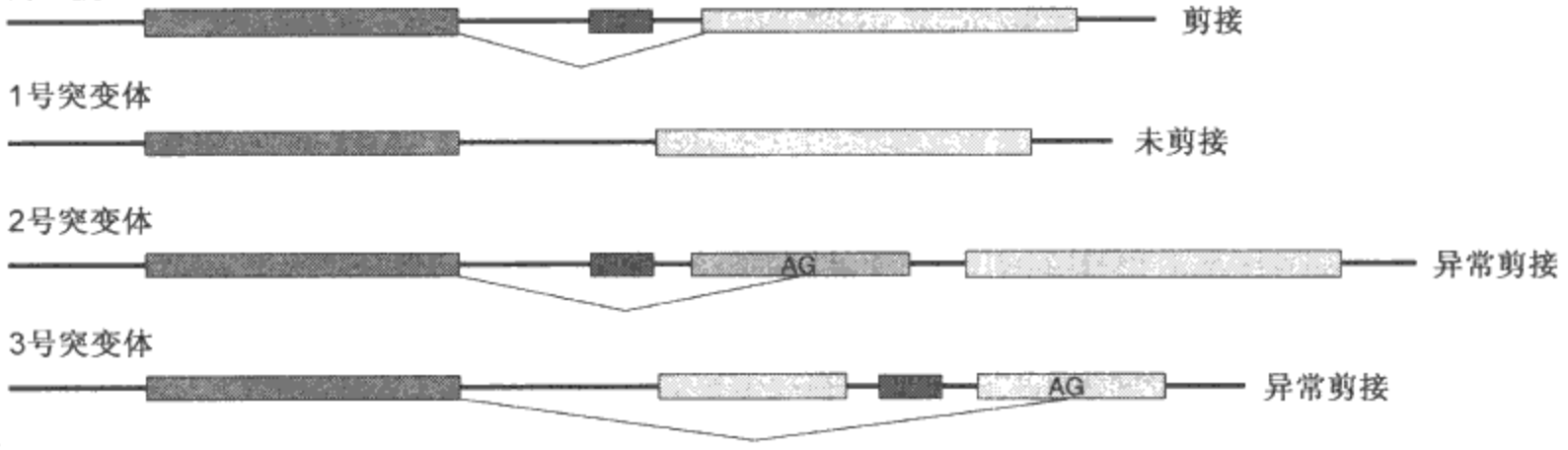


图 14.8　酵母基因内含子中关键剪接信号的证实。Langford 和 Gallwitz 在体外突变酵母的 *actin* 基因，并转入正常酵母细胞，然后检测这些突变基因的剪接情况。野生型基因含有两个外显子（蓝色和黄色），所有内含子均有一保守序列（红色），野生型基因在酵母细胞内能够被正确剪接。Langford 和 Gallwitz 将这一内含子保守序列缺失而获得突变体＃1，该突变基因的转录物不能被剪接。突变体＃2 是将一额外的 DNA 片段（紫色）插入到内含子保守序列的下游，此突变基因的转录物会在插入序列的第一个 AG 处发生异常剪接。突变体＃3 是将内含子的保守序列插入到下游的第二个外显子内，此突变基因的转录物会在内含子保守序列下游的第一个 AG 处发生异常剪接。这些结果表明，内含子的保守序列在 RNA 剪接过程中发挥重要作用，它能够将其下游的 AG 确定为 3′-剪接位点。

1983 年，Christopher Langford 和 Dieter Gallwitz 在对酵母肌动蛋白（actin）基因的研究中，首次为内含子内存在保守分支点序列提供了线索。研究者对克隆的肌动蛋白基因进行突变，产生大量突变基因，然后用这些突变基因转化正常的酵母细胞，通过 S1 作图分析剪接情况。图 14.8 显示了实验结果。首先，当位于内含子 3′-剪接位点上游 35～70bp 间的序列缺失（突变体＃1）后，剪接被阻断，表明长度为 35bp 的缺失区域含有一个对剪接十分重要的序列，如图 14.8 中的红色小条框所示。其次，当在这个特殊序列与第二外显子之间插入额外的 DNA 片段（突变体＃2）时，5′-剪接位点可正常剪接，3′

端也可剪接，但剪接并没有在正确的 3′-剪接位点处进行，而是在异常的 3′-剪接位点，即特殊内含子序列下游的第一个 AG 处发生剪接，该 AG 位于插入的 DNA 片段内。这一结果表明，该特殊内含子序列的作用是确定剪接位点，指导剪接机构在其下游适当距离的 AG 处进行剪接。如果在正常 AG 剪接位点的前面插入一新的 AG 序列，那么剪接可能就会在新位点处进行。最后，突变体 #3 在其第二个外显子内含有这个特殊的内含子。果然，在这种情况下，特殊内含子下游的第一个 AG 位点便成了 3′-剪接位点，而且该 AG 恰好位于第二个外显子内。

此特殊内含子序列的重要性在于该序列含有分子点腺嘌呤核苷酸，即 UACUAAC 序列中最后的 A。事实上，所有酵母细胞核内含子都含有这个几乎不变的保守序列，但高等真核生物的分支点序列并没有很强的保守性，其分支点 A 的周围序列为 $U_{47}NC_{63}U_{53}R_{72}\underline{A}_{91}C_{47}$，其中 R 表示任一嘌呤碱基（A 或 G）；N 表示任一碱基，下标数字表示该碱基在该位置出现的频率。例如，分子点 A（具下画线）在其位置处出现的概率为 91%。序列中的第一个 U 经常会被 C 所取代，所以该位点通常为嘧啶碱基。

小结 细胞核基因的内含子除了在其 5′端和 3′端具有保守序列外，在分支点处也存在着分支点保守序列。在酵母中，固有的分支点序列为 UACUAAC。高等真核生物的分支点序列保守性不强，但不管怎样，分支点核苷酸是该序列中的最后一个 A。

剪接体

1985 年，Edward Brody 和 John Abelson 通过实验发现，酵母的套索状剪接中间体并非游离于溶液之中，而是与沉降系数为 40S 的剪接体（spliceosome）结合在一起。研究人员将标记的前体 mRNA（pre-mRNA）加入到无细胞提取物中，利用甘油梯度超速离心分离纯化剪接体。图 14.9 显示显著的 40S 复合体峰中含有标记 RNA。对这些 RNA 的凝胶电泳分析显示，剪接中间体和切离的内含子具有套索结构。为验证剪接体在剪接中的重要性，Brody 和 Abelson 将 pre-mRNA 分支位点 A 突变为 C，试图利用该突变 mRNA 形成剪接体，结果其形成剪接复合体的能力被极大地削弱。Sharp 及同事于 1985 年从人的 HeLa 细胞中分离出了沉降系数为 60S 的剪接体。

剪接体显然含有 pre-RNA，但还含有许多其他 RNA 分子和蛋白质。其中有些 RNA 和蛋白质结合在一起形成**核内小核糖核蛋白体**（small nuclear ribonuclearprotein，**snRNP**，发音为“snurps”），是

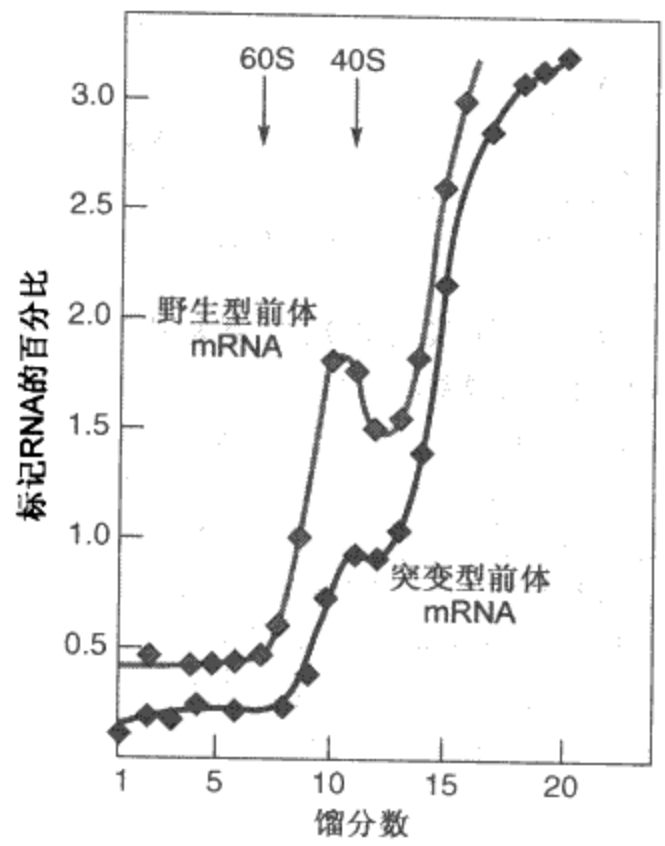


图 14.9 酵母剪接体。Brody 和 Abelson 将标记的酵母 pre-mRNA 与剪接提取物共同孵育，然后进行甘油梯度超速离心，最后通过液体闪烁计测定每一组分的放射性强度。图中显示了分别用野生型 pre-mRNA（红色）及在 5′ 剪接点发生点突变的 pre-mRNA（蓝色）进行实验获得的结果，野生型 pre-mRNA 明显与 40S 的聚合体结合在一起，而突变的 pre-mRNA 与之结合能力要弱得多。(*Source*: Adapted from Brody, E. and J. Abelson, The spliceosome: Yeast premessenger RNA associated with a 40S complex in a splicing-dependent reaction. *Science* 228: 965, 1985.)

由偶联到蛋白质上的**核内小 RNA**（small nuclear RNA，**sn RNA**）组成。通过凝胶电泳可将 snRNA 分离开来，共包括 5 类：Ul、U2、U4、U5 和 U6，所有这 5 类 RNA 均参与了剪接体的形成，并在剪接中发挥着关键作用。

理论上讲，在内含子末端和分支点处的保守序列可被蛋白质或核酸所识别，我们现在已有充分的证据证明 snRNA 和蛋白质剪接因子是识别关键剪接信号的重要分子。图 14.10 是两侧连接外显子的典型内含子结构，以及在关键位点发生相互作用的分子。在本章的后面内容中，我们将详细介绍与这些相互作用有关的证据。

小结 RNA 的剪接是在剪接体中进行的，酵母剪接体和哺乳动物剪接体的沉降系数大体分别为 40S 和 60S，剪接体中包含有 pre-RNA、snRNP 以及能够识别剪接信号的蛋白质剪接因子，协同完成剪接。

U1 snRNP 1980 年，Joan Steitz 以及 J. Rogers

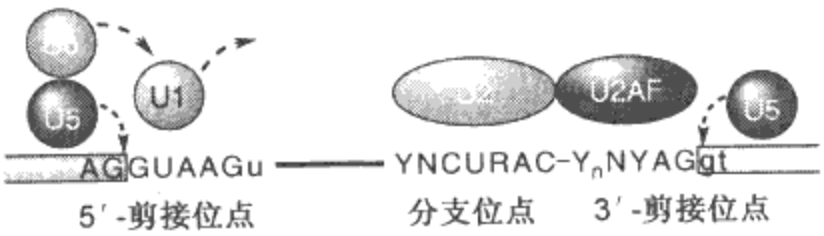


图 14.10　RNA 及蛋白质对典型哺乳动物 pre-RNA 内含子的识别。大写字母表示该位点的碱基具有较好的保守性，小写字母表示该位点碱基的保守性较差。Y 代表嘧啶碱基，R 代表嘌呤碱基，N 代表任意碱基。首先由 U1 snRNP 识别 5′-剪接位点，随后被 U6 snRNP 取代，U2 snRNP 识别分支位点，U2AF 蛋白（U2 相关因子）识别 3′-剪接位点。其他因子完成最初的识别之后，U5 snRNP 结合在 5′-剪接位点和 3′-剪接位点上。

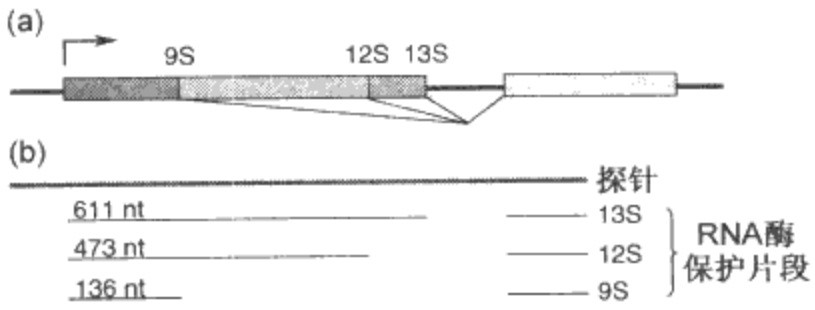


图 14.11　腺病毒 *E1A* 基因的剪接示意图及剪接产物的 RNase 保护分析。（a）剪接示意图。3 个可变 5′-剪接位点（分别位于红色、橙色、蓝色长条的边界处及蓝色长条的末端）和单一的 3′-剪接位点（位于黄色长条的起始处）发生剪接，分别产生 9S、12S、13S 的 mRNA 分子。（b）RNase 保护实验分析。紫色线条代表标记的 RNA 探针，经 RNase 处理后，被探针保护的片段即为剪接的产物，其长度各不相同［在对应片段的上方标出了片段的核苷酸长度（nt），3 个剪接产物均产生了对应于下游外显子的保护片段］。（*Source*：Adapted from Zhuang，Y. and A. M. Weiner，A compensatory base change in U1 snRNA suppresses a 5′-splice site mutation. *Cell* 46：829，1986.）

和 R. Wall 分别注意到，U1 snRNA 具有一段几乎能够与位于 5′-和 3′-剪接位点保守序列完全互补的序列。研究者提出：U1 snRNA 通过碱基互补配对而将两个剪接位点联系在一起，便于剪接反应的进行。我们现在知道，pre-RNA 的剪接还涉及内含子的分支位点，但在这个简化的机制中被忽略了。不过，U1 snRNA 与 5′-剪接位点发生碱基配对是剪接反应发生的关键。

1986 年，Yuan Zhuang 和 Alan Weiner 的遗传学实验证明了 5′-和 3′-剪接位点与 U1 碱基配对的重要性。研究者对腺病毒 *E1A* 基因内 3 个可变 5′-剪接位点中的一个进行了突变。*E1A* 基因的 3 个 5′-剪接位点均能与共同的 3′-剪接位点完成对前体转录物的正常剪接，产生 9S、12S 和 13S3 种成熟 mRNA 分子（图 14.11）。发生在 12S 5′-剪接位点的突变可破坏该位点与 U1 间的碱基配对。为检测突变对剪接的影响，Zhuang 和 Weiner 从转化有 5′-剪接位点发生突变的 *E1A* 基因的细胞中分离 RNA，然后进行 RNase 保护实验（第 5 章）。图 14.11 显示了信号核苷酸序列的长度（nt），与预期分别从 3 个 5′-剪接位点起始剪接获得产物的长度相一致。

Zhuang 和 Weiner 测试的第一个突变体实质是个双突变体。将内含子的第 5 个和第 6 个碱基（+5 和+6）GG 突变成 AU（图 14.12）时，破坏了内含子第 5 个碱基 G（+5）与 U1 snRNA 的碱基 C 之间的配对，但使内含子第 6 个碱基 U（+6）与 U1 snRNA 的碱基 A 形成新的碱基对，尽管突变并没有造成配对碱基数目的改变，但配对碱基间的结合力受到了削弱，因为连续碱基对的数目减少了，剪接会受到这种突变的影响吗？图 14.13（泳道 4）显示，突变基本上中断了 12S 位点的剪接，却促进了 13S 和 9S 位点的剪接。接着，研究者在 U1 snRNA 内引入一个补偿突变，使之能够恢复与突变剪接位点的配对。然后将突变的 U1 基因用相同的质粒携带导入 HeLa 细胞。图 14.13（泳道 5）显示，突变的 U1 不仅能够恢复与剪接位点的碱基配对，而且剪接又可在 12S 位点正常进行。

由此可见，剪接位点与 U1 snRNA 间的碱基配对是进行正常剪接所必需的。但是，这种碱基配对是否足以确保正常剪接的发生？如果在剪接位点内引入突变，减弱剪接位点与 U1 的配对能力，而相应的剪接不会被 U1 的补偿突变所抑制，则可说明这个碱基对不足以确保剪接的发生。图 14.12 和图 14.13 示意了 Zhuang 和 Weiner 所开展证明实验的过程。研究者将 13S 5′-剪接位点中+3 位的 A 突变为 U，使原本 6 个连续的碱基对被间断，阻断了 13S 剪接事件，但却促进了 12S 剪接，同时在某种程度上也促进了 9S 剪接（图 14.13，泳道 6）。对 U1 基因的补偿突变可恢复连续的 6 个碱基对，但不能恢复在 13S 位点的剪接（泳道 7）。由此可见，5′-剪接位点与 U1 snRNA 间的碱基互补配对并不是剪接的充分条件。

小结　遗传学实验表明，U1 snRNA 与 mRNA 前体 5′-剪接位点的互补配对是进行正常剪接所必需的，但并不是剪接发生的充分条件。

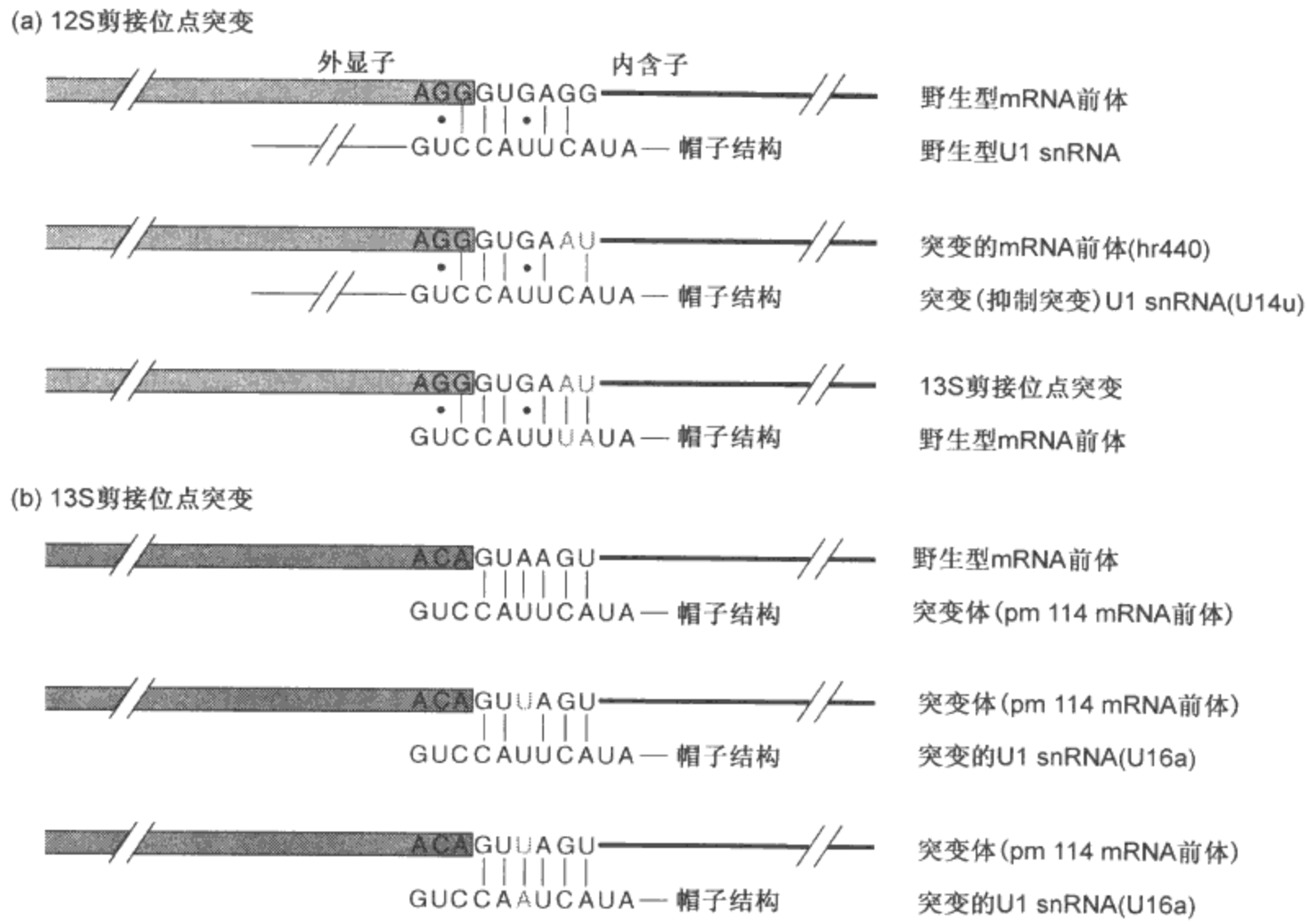


图 14.12　野生型和突变型 5′-剪接位点与野生型和突变型 U1 snRNA 的互补序列图。（a）12S 剪接位点突变。图右侧标注出野生型和突变型序列，mRNA 前体和 U1 RNA 之间的 Watson-Crick 碱基配对用竖线表示，圆点表示碱基配对的摇摆性。红色字母表示突变的碱基。如图 14.11 那样用橙色框表示外显子的末端。（b）13S 剪接位点突变。所有的符号表与（a）相同，只是如图 14.11 那样用蓝色框表示外显子的末端。

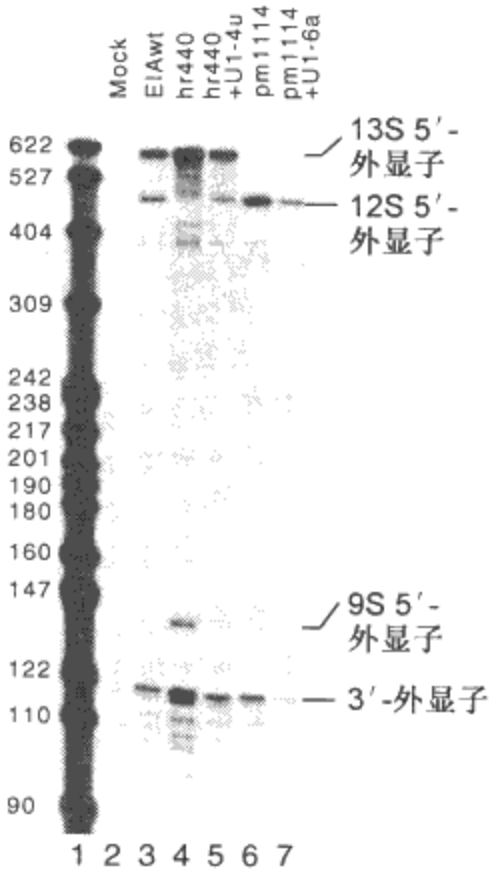


图 14.13　RNase 保护分析实验结果。Zhuang 和 Weiner 将含有如图 14.12 所示的野生型和突变型 5′-剪接位点的基因，以及野生型和突变型 U1 snRNA 基因转化 HeLa 细胞。然后利用图 14.14 示意的 RNase 保护实验分析剪接情况。泳道 1 为分子质量标准，左侧数字表示片段的碱基对长度；泳道 2 为非转化细胞（阴性对照）；泳道 3 为野生型 *E1A* 基因及野生型 U1 snRNA，能够检测到 12S 和 13S 产物，但没有检测到 9S 产物，因为 9S 产物只有在转染的后期才能产生；泳道 4 为 12S 5′-剪接位点发生改变的突变基因 *hr440*，无 12S 产物的形成；泳道 5 为突变体 *hr440*＋突变体 U1 snRNA（U1-4u），12S 的 5′-位点的剪接恢复；泳道 6 为 13S 5′-剪接位点发生改变的突变基因 *pm1114*，无 13S 产物的形成；泳道 7 为突变体 *pm1114*＋突变的 U1 snRNA（U1-6a），虽然恢复了 5′-剪接位点与 U1 snRNA 之间的碱基配对，但并没有发生 13S 剪接事件。（*Source*：Zhuang Y. and A. M. Weiner, A compensatory base change in U1 snRNA suppresses a 5′-splice site mutation. *Cell* 46（12 Sept 1986）f. 1a, p. 829. Reprinted by permission of Elsevier Science.）

U6 snRNP　为什么 U1 碱基的改变有时并不能　补偿由 5′-剪接位点突变所造成的剪接抑制？对于这

个问题，我们可能会想象出不同的答案，包括可能有某些蛋白质或蛋白体参与对 5′-剪接位点序列的识别。在这种情况下，U1 碱基的改变并不能恢复剪接体对 5′-剪接序列的识别。也可能有其他 snRNA 参与了与 5′-剪接位点序列的碱基配对。改变 U1 的碱基序列使之恢复与突变剪接位点的配对，但没有恢复突变剪接位点与其他 snRNA 之间的碱基配对，因此剪接反应仍处于抑制状态。

Christine Guthrie 小组与 Joan Steitz 小组证实，确实有其他 snRNA 与 5′-剪接位点序列进行碱基配对。这个 snRNA 分子就是 U6 snRNA。Steitz 发现 U6 可以与内含子+5 位碱基发生生化学交联，从而首次探明 U6 参与发生在 5′-剪接位点附近的分子生物学事件。基于这个发现，Steitz 认为 U6 snRNA 固有序列 ACAGAG 中的 ACA 能够与内含子 5′-剪接位点保守的+4～+6 位碱基 UGU 形成稳定的碱基对（图 14.14）。

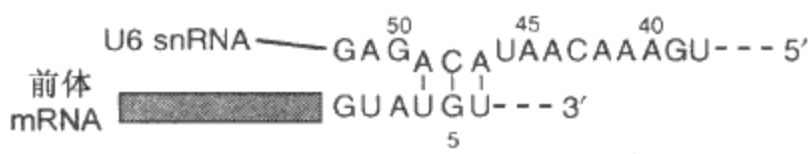


图 14.14　酵母 5′-剪接位点与 U6 snRNA 相互作用模型。U6 snRNA 保守基的 ACA（第 47～49 nt）与内含子的第 4～6 个碱基 UGU 形成稳定的碱基对。（*Source*：Adapted from Lesser, C. f. and C. Guthrie, Mutations in U6 snRNA that alter splice site specificity: implications for the active site. *Science*: 262: 1983, 1993.）

Erik Sontheimer 和 Joan Steitz 通过化学交联研究显示，剪接体中 U6 的结合位点十分接近于内含子的 5′-剪接位点。实验策略如下，首先合成具有两个外显子和一个内含子的模式剪接前体（腺病毒主要晚期基因座的前两个外显子和第一个内含子），然后用 4-硫尿嘧啶（4-thioU）取代第一个外显子的最后一个核苷酸或内含子的第二个核苷酸，光敏感的 4-thioU 被紫外线激活后，就会与其所接触的 RNA 分子发生共价交联。通过对交联分子的结构分析，他们发现了能够与 5′-剪接位点互补配对的 RNA 分子。

图 14.15 显示，当 Sontheimer 和 Steitz 用 4-thioU 取代内含子的第二个核苷酸后，模式剪接前体能够与其他 RNA 分子发生交联。除了与 U2 交联外（相关内容在本章后面讨论），U6 也能与内含子的第二个核苷酸发生交联。而且，通过分析还揭示了 U6-内含子交联体形成的时间，即在剪接反应进行到 20min 后 U6-内含子交联体才开始大量出现，相比之下，内含子-外显子 2（内含子-E2）套索中间体的数量在剪接反应开始 15min 后达到峰值。最初，这一发现似乎暗示在 U6 与内含子 5′端结合之前套索分子的形成是剪接反应的第一步。然而，正如我们将要了解到的那样，已有充分的证据表明，在剪接反应的第一步发生之前，U6 就已经与 5′-剪接位点相结合了。因此，有可能在剪接的起始步骤，U6 就已经与内含子的特殊核苷酸结合了。

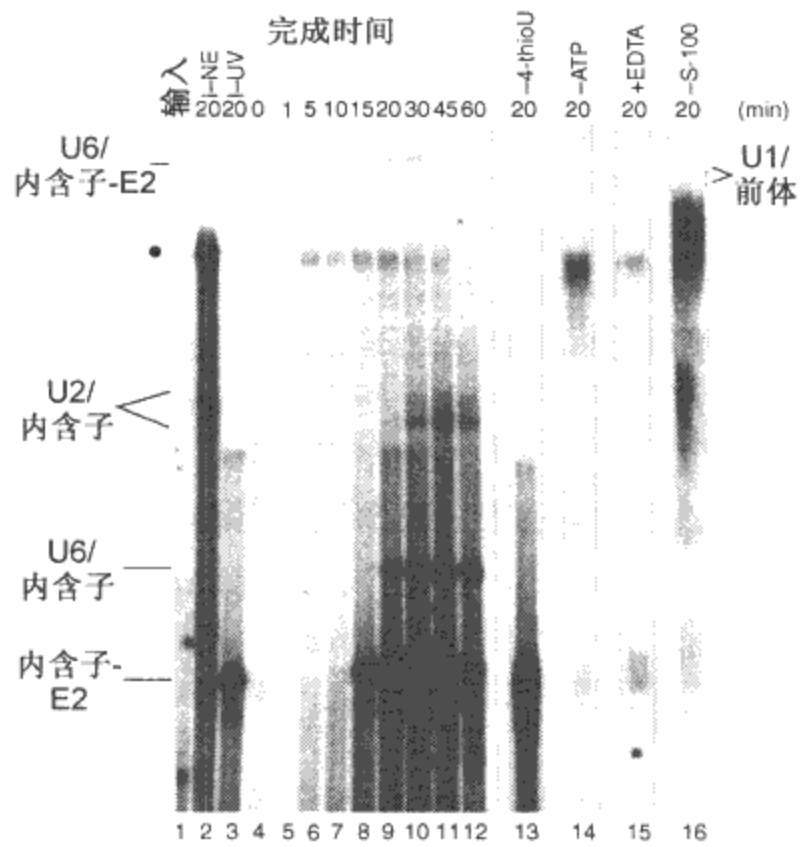


图 14.15　剪接中间体与 4-硫尿嘧啶之间的交联。Sontheimer 和 Steitz 将内含子第二个核苷酸为 4-thioU 的标记腺病毒剪接前体与细胞核剪接提取物混合孵育，在不同时间段内（图上方所示）分离样品，经紫外线照射，使 4-thioU 与邻近的 RNA 交联，将产物进行凝胶电泳。泳道 1：未孵育的 RNA；泳道 2：在无细胞核提取物（NE）存在条件下温育 20min；泳道 3：孵育 20min，但样品不经紫外线照射；泳道 4～12：孵育不同的时间；泳道 13：无 4-thioU 标记；泳道 14：无 ATP；泳道 15：加入 EDTA 螯合 Mg^{2+}，阻断剪接；泳道 16：经高速离心纯化得到的组分，取代细胞核提取物。左侧圆点指示分子内交联产生的条带；E2 代表外显子 2。（*Source*：Sontheimer E. J. and J. A. Steitz, The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science* 262 (24 Dec 1933) f. 3, p. 1991. Copyright © American Association for the Advancement of Science.）

如何确知在套索状中间体形成之前 U6 已完成了与剪接底物的结合？David Wassarman 和 Joan Steitz 用另一种交联实验和 Northern 印迹实验证明

了这一点。与直接在剪接底物中掺入交联试剂不同，研究者将交联试剂［补骨脂素（psoralen）］添加到溶液中，该实验方案的缺点是交联试剂不能进入剪接体的活性位点，无法显示形成的 RNA-RNA 复合物，但也有其优点，那就是能够在剪接反应的任何时间内，显示与剪接底物相结合的 RNA 种类。

图 14.16 显示了补骨脂素交联实验的结果，证明 U6 与剪接底物之间的结合（以及其他 RNA-RNA 的结合）。Wassarman 和 Steitz 采取的实验步骤如下。

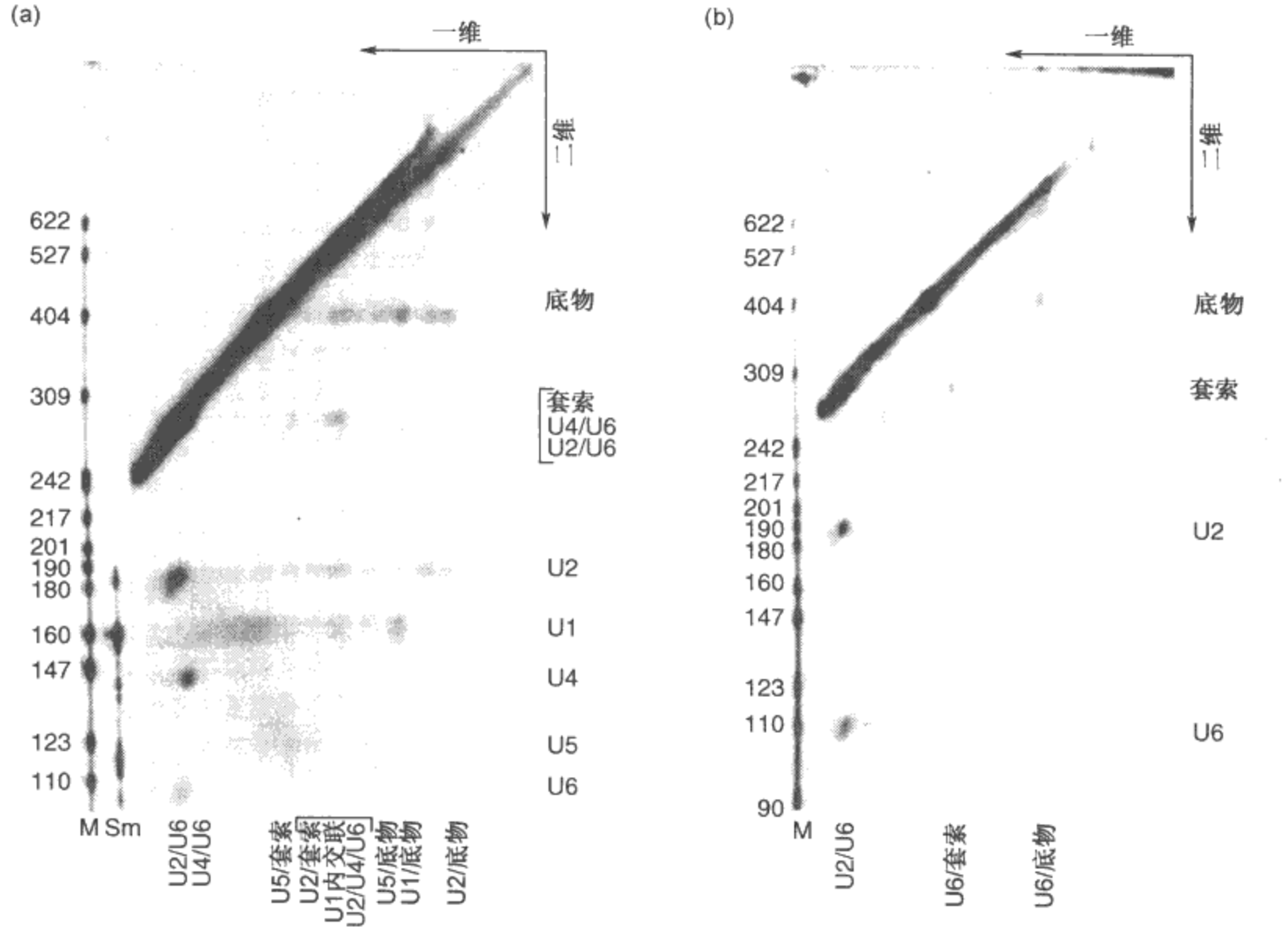


图 14.16　剪接复合体的二维电泳检测。（a）含有 U1、U2、U4 和 U5 的复合体。Wassarman 和 Steitz 利用生物素标记的剪接前体形成剪接复合体，用补骨脂素交联复合体中的 RN 子。用 m_3G 抗体（识别除 U6 以外的 snRNPA）沉淀含 snRNP 的复合体，然后用链霉亲和素磁珠分离含剪接前体的复合物。接下来，用［^{32}P］pCp 和 RNA 连接酶对复合体中 RNA 进行末端标记，随后对复合物进行水平一维凝胶电泳，在 254nm 紫外线条件下解离复合体，然后进行垂直二维凝胶电泳（如图上方所示）。图右侧标出分子标准的位置，以此鉴定复合体的 RNA。图下方标明了复合体中的混合 RNA。有时，两个或更多的复合体会共纯化，如 U2/U4/U6、U1 的内交联分子、U2-套索复合体。（b）含 U6 的复合体。由于 U6 无 m_3G 帽子结构，Wassarman 和 Steitz 利用生物素标记的 U6 序列特异性寡核苷酸和链霉亲和素分离含 U6-snRNP 的复合体，其他情况同图（a）中所述。图右侧标出分子标准的位置，以此鉴定含有剪接底物、套索分子、U2 的 U6 复合体，在前两个复合物中并没有检测到 U6 的存在，但 U6 确实存在，因为这两个复合体是利用 U6 特异性探针筛选到的。Sm 是指能够与来自红斑狼疮患者的抗 snRNP 的抗血清发生免疫共沉淀反应的 snRNA，这种抗血清能够与存在于 U1、U2、U4、U5 和 U6 snRNP 中的一组蛋白质（Sm 蛋白）发生反应；Sub 是指剪接底物。（*Source*：Wasserman，D. A. and J. A. Steliz. Interactions of small nuclear RNA's with precusor messenger RNA during in vitro splicing. *Science* 257（25 Sept 1992）f. 2，p. 1919. Copyright © American Association for the Advancement of Science.）

1. 将用生物素标记的腺病毒主要晚期剪接底物与 HeLa 细胞的核提取物共同孵育 30min，使腺病毒剪接底物易于被纯化。

2. 向混合物中加入补骨脂素，并在 365nm 光照

条件下继续孵育 60min，使剪接复合体内的 RNA 发生交联。

3. 通过双步选择，收集所有剪接复合体。按照定义，这些复合体应该包含两种物质：snRNP 和剪接前体或套索状中间体。对 snRNP 进行二步选择，第一步是分离含有 snRNP（Ul、U2、U4、U5 和 U6）的复合体。除了 U6 之外，所有其他 snRNP 均具有三甲基鸟嘌呤（m_3G）结构，因此可通过能与 m_3G 发生免疫沉淀反应的抗体而有选择地筛选复合体。为得到含 U6 的复合体，研究者利用生物素标记的 U6 特异性寡核苷酸，所有含 U6 的复合体会与该寡核苷酸杂交，又因为链霉亲和素（streptavidin）可与生物素紧密结合，含 U6 的复合体与链霉亲和素磁珠一起形成沉淀。第二步，Wassarman 和 Steitz 利用链霉亲和素磁珠，从第一步获得的复合体中筛选具生物素标记剪接底物复合体。这种方法不能用于含 U6 复合体的筛选，因为 U6 复合体中无生物素标记的剪接前体。

4. 用 RNA 连接酶将［5′-^{32}P］pCp 连接到 RNA 分子的 3′端，从而标记所选复合体中的 RNA。

5. 将筛选出的并经标记处理的复合体进行一维凝胶电泳。

6. 电泳结束后，将凝胶于 254nm 光照条件下照射 20min，破坏复合体中 RNA 分子间的交联。

7. 将凝胶调转 90°后，进行二维凝胶电泳。

8. 将凝胶进行放射自显影，检测标记的 RNA。

交联逆转的优点非常明显，不管是存在于复合体中还是以游离状态单独存在，RNA 分子的电泳迁移率不会发生改变，放射自显影后，这些 RNA 分子在凝胶上应该排列成一条对角线。然而，因交联逆转导致复合体分离并产生游离的 RNA，这些单体 RNA 分子在二维电泳时泳动速率加快，出现在对角线的下方，从而使之与复合体分开。因此可以通过迁移率鉴定这些 RNA。图 14.16 显示的复合体其形成涉及 Ul、U2、U4、U5、剪接底物以及套索状中间体。U6 能与剪接底物、套索状中间体形成复合体，在图 14.16 中很难发现 U6 的存在，但底物和套索状中间体可清晰地显现于图中。由于复合体是通过 U6 特异性寡核苷酸筛选出来的，所以我们可认定 U6 的存在。由此可见，在剪接的起始步骤前后，U6 均与剪接底物结合在一起。由于 U2 snRNA 与 U6 snRNA 具有互补配对的碱基序列，推测 U2 与 U6 可形成复合体，本实验也证实了 U2-U6 复合体的存在。U2 与 U6 间的碱基配对有助于剪接体活性位点结构的形成，相关内容将在本章后面介绍。以标记的反义 snRNA 作探针，通过 Norther 印迹鉴定了每个复合体中 snRNP 的种类。用 U6 的反义 RNA 为探针检测到了 U6-底物和 U6-套索复合体的存在，并表明 U6-套索复合体是在剪接反应开始的 15min 内形成的。

小结 通过 U6 snRNA 与内含子 5′-剪接位点间的碱基配对，U6 snRNP 可结合在内含子的 5′端。在形成套索状中间体之前，U6 snRNP 已完成与内含子的结合，但剪接的第一步完成后，其性质可能发生了改变。U6 与剪接底物的结合是进行正常剪接反应所必需的，在剪接过程中，U6 也可能与 U2 结合。

U2 snRNP 如图 14.17 所示，酵母分支点保守一致性序列与 U2 snRNA 的序列互补，遗传学分析表明两者之间的碱基配对是剪接所必需的。Christine Guthrie 及同事为这一论点提供了遗传学证据，研究者通过突变分析证实，分支位点突变可抑制剪接，而且这种抑制效应可被 *U2* 基因的补偿突变所逆转。

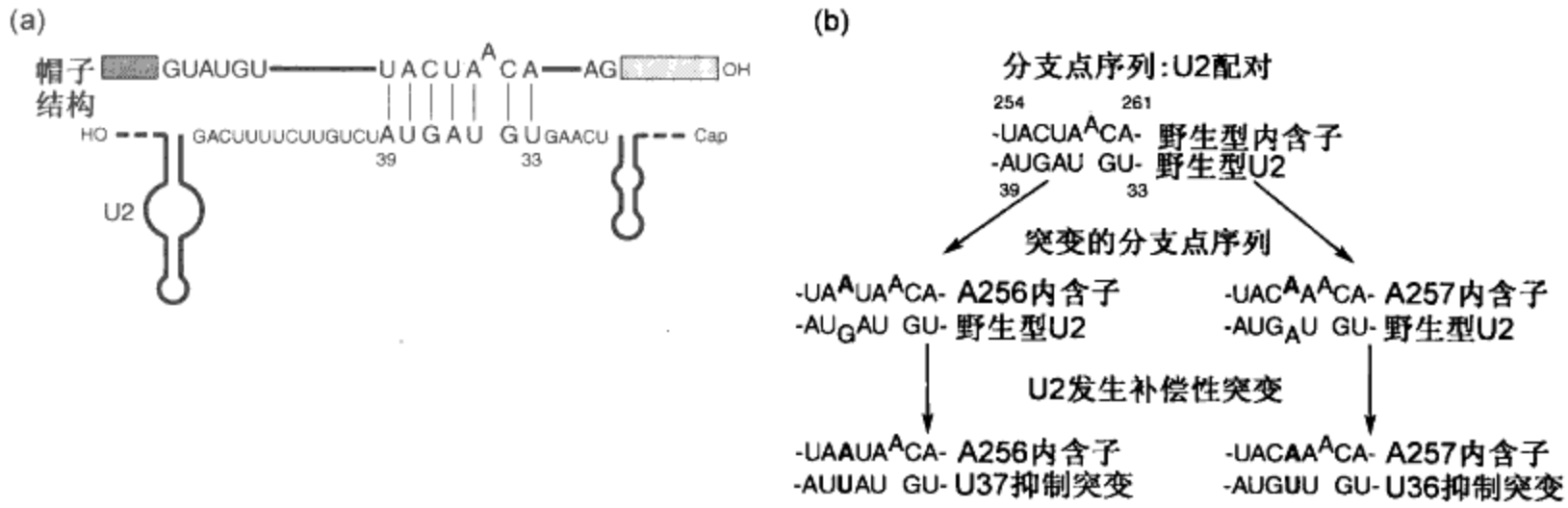


图 14.17 酵母 U2 与分支位点序列的碱基配对。（a）推测的野生型酵母 U2 与分支点不变序列间的碱基配对。注意，突出的分支点 A（上方）不参与碱基配对。（b）推测的野生型和突变型 U2 及分支点间的碱基配对。红色字母（A）表示分别在分支位点的第 256 位和第 257 位碱基处引入的突变碱基；绿字母（U）表示在 U2 内引入的补偿性突变。(*Source*: Adapted from Parker R., P. G. Sliciano, and C. Guthrie, Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U_2-like snRNA. *Cell* 49: 230, 1987.)

为开展相关的实验工作，研究者利用带有 *actin-HIS4* 融合基因的组氨酸依赖型酵母突变菌株，其中 *actin* 部分包含一个内含子，如果这个融合基因的转录物被正确剪接，则融合基因的 *HIS4* 部分就能编码产生有活性的蛋白质，可以将组氨酸前体组氨醇转化为组氨酸，该突变菌株就可在含组氨醇的培养基上存活。接着，研究者在剪接分支位点处引入突变。其中之一是将第 257 位碱基 U 突变为 A，即分支点保守一致性序列由 UACUAAC 突变为 UACAAAC，则 95%的剪接事件被抑制，该突变菌株在含组氨醇的培养基上不能存活；另一个突变是将第 256 位碱基 C 突变为 A，即分支位点保守一致性序列由 UACUAAC 突变为 UAAUAAC，则 50%的剪接事件被抑制。

为检测 U2 突变对分支点序列突变的抑制效应，Guthrie 及同事将携带 *U2* 突变基因的质粒转入酵母细胞。为筛选转化细胞，研究者使质粒带有选择标记 *LEU2* 基因（宿主细胞为 *LEU*⁻）。由于酵母突变细胞的 *U2* 基因只有一个拷贝，突变后会导致其他所有基因的剪接无法正常进行，所以还需为转化细胞提供一个额外的 *U2* 基因拷贝。图 14.18 显示，突变的 U2 恢复了与突变分支点的碱基配对，从而恢复了正常的剪接反应。这种情况在 A257 突变体内尤为明显，在野生型 U2 存在条件下没有观察到该突变体的生长，但在突变型 U2 存在条件下，U2 恢复了与突变分支点的碱基配对，该突变菌株能够快速生长。

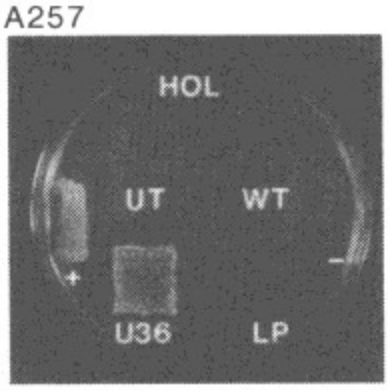


图 14.18　通过突变抑制证明 U2 snRNP 与分支点的碱基配对。将 A257（a）和 A256（b）突变体培养在 HOL 培养基上，在野生型菌株和 U2 抑制突变菌株存在的情况下，衡量突变菌株在培养基上的生长状况。每一细胞菌落下方的缩写标明了细胞内 U2 的性质。UT：未转化（无 *U2* 基因导入）；WT：野生型 *U2* 基因；U36：能够恢复与 A257 位碱基配对的 *U2* 突变；U37：能够恢复与 A256 位碱基配对的 *U2* 突变；LP：丢失 U2 质粒的菌落。每个平皿的（＋）表示正对照菌株，含有野生型融合基因，无额外的 *U2* 基因拷贝；每个平皿的（－）表示阴性对照菌株，无融合基因。（*Source*：Parker R.，P. G. Siciliano，and C. Guthrie，Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell* 49（24 Apr 1987）f. 3，p. 232. Reprinted by permission of Elsevier Science.）

除了能与分支点进行碱基配对外，U2 也能与 U6 进行碱基配对。基于这两类 RNA 的碱基序列可推测两者间会发生结合，Guthrie 及同事通过遗传学分析为 U2 与 U6 间的碱基配对提供了直接证据。首先，Guthrie 及同事发现，酵母 U6 的 ACG 序列突变会产生致死表型，该序列能与 U4 snRNA 发生碱基配对。研究者通过两种途径证实，这些突变体干扰 U6 snRNA 与 U4 碱基配对并不是致死表型产生的原因。首先，研究者在 U4 内引入能够破坏 U4 与 U6 之间碱基配对的突变，并不影响细胞的正常生长；其次，但当 U6 发生致死突变后，即使在 U4 中引入能够恢复与突变 U6 碱基配对的补偿突变，也不能抑制 U6 的致死突变效应。

很显然，除了干扰与 U4 间的相互作用外，U6 发生的致死突变还干扰了 U6 与其他分子间的相互作用。Hiten Madhani 和 Christine Guthrie 证明，U2 是另一个能够与 U6 相互作用的分子。在 U6 的第 56～59 位碱基引入致死突变，而这些突变可以被 U2 的第 23 位、第 26～28 位碱基的补偿突变所抑制，因为突变了的 U2 可以恢复与突变 U6 间的碱基配对。U2 与 U6 间的关键碱基配对区称为螺旋 I，图 14.22 总结了这一结构的特点。

其他研究者（Jian Wu 和 James Manley 以及 Ban-shidar Datta 和 Alan Weiner）利用相同的遗传分析策略，研究了哺乳动物细胞的剪接效率。结果表明，U2 的 5′端与 U6 的 3′端也能发生相互作用，形成的碱基配对区称为螺旋 II。U2 突变能够被 U6 恢复碱基配对的补偿突变所抑制。这种相互作用在酵母是非必需的，但对哺乳动物是必需的，至少可提高剪接效率。

小结　U2 snRNA 能够与剪接分支位点的保守序列进行碱基互补配对。这种碱基配对是剪接反应

所必需的。U2 与 U6 形成的关键碱基对构成了螺旋 I 结构区，对参与剪接的 snRNP 具有定位功能。此外，U2 的 5′端与 U6 的 3′端也能发生相互作用，形成的碱基配对区称为螺旋 II，这一区域对于哺乳动物细胞的剪接十分重要，但对酵母细胞而言则是非必需的。

U5 snRNP 我们已经了解了有关 U1、U2 和 U6 snRNP 参与剪接的证据。那么，U5 的作用如何？U5 与其他 snRNP 或剪接底物的保守序列无明显互补性，但 U5 可能与两个外显子相结合，在剪接的第二步将外显子放置在适当位置处。

Sontheimer 和 Steitz 用 4-thioU 取代剪接底物碱基的方法，为剪接过程中 U5 与两个外显子末端发生相互作用提供了证据。在其中一个实验中，研究者用 4-thioU 取代腺病毒主要晚期剪接底物的第二个内含子的第一个碱基 C，剪接底物仍能进行正常剪接。将 4-thioU 与靠近第二个外显子 5′端的任一 snRNA 交联，在剪接进行 30min 后，可分离获得一个二联体复合物，即 U5/内含子-E2（图 14.19）。此时第一步剪接反应早已完成，许多其他的复合体也已形成，但在这里对这些复合体并不做讨论。

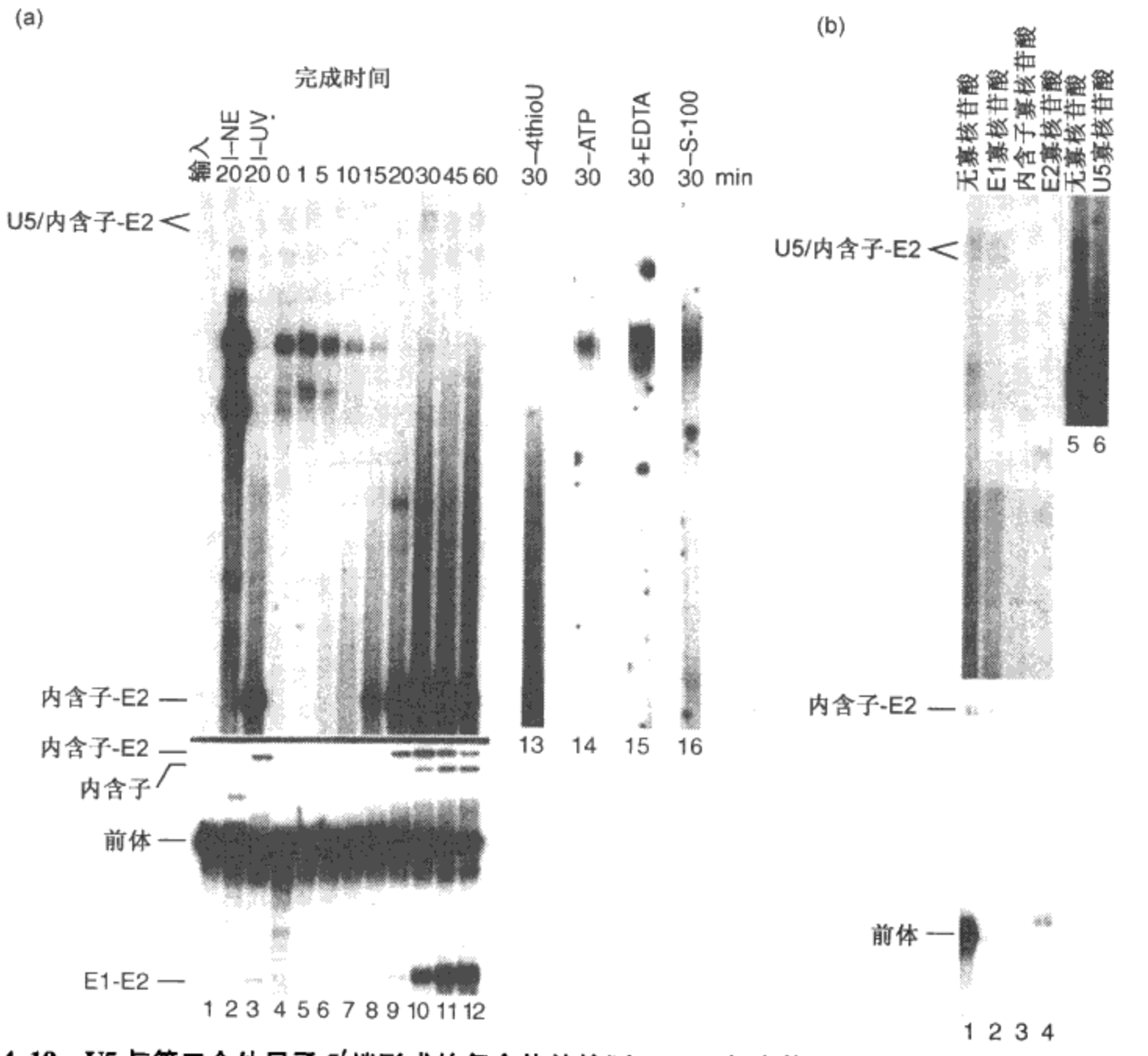


图 14.19 U5 与第二个外显子 5′端形成的复合体的检测。（a）复合体的形成。Sontheimer 和 Steitz 将 4-thioU 引入到标记剪接底物的第二个外显子的第一个碱基处，然后，在剪接过程的不同时间段内，使之与邻近的 RNA 交联。电泳交联产物并放射自显影检测。位于凝胶上方的双联体 U5/内含子-E2 产生于剪接过程的后期（约 30min 后），泳道的标注同图 14.15。（b）复合物中 RNA 的鉴定。Sontheimer 和 Steitz 在剪接开始 30min 后照射剪接混合物，使之形成交联物。将能够与 U5 及其他 RNA 分子互补的 DNA 寡核苷酸与交联物共同孵育，然后加入 RNase H，降解与寡核苷酸杂交的 RNA，电泳后放射自显影检测。泳道 1 和 5：无寡核苷酸；泳道 2：外显子 1 反义寡核苷酸；泳道 3：内含子反义寡核苷酸；泳道 4：外显子 2 反义寡核苷酸；泳道 6：U5 反义寡核苷酸。内含子反义寡核苷酸、外显子 2 反义寡核苷酸、U5 反义寡核苷酸均有助于复合体解离，表明此复合体是由内含子、第二个外显子及 U5 组成。(*Source*: Sontheimer E. J. and J. A. Steitz. The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science* 262(24 Dec 1993)f. 4, p. 1992. Copyright © American Association for the Advancement of Science.)

为验证这个二联体复合体确实包含 U5、内含子和 E2 等组分，Sontheimer 和 Steitz 利用能与这些 RNA 互补的 DNA 寡核苷酸对复合体进行杂交，然后用 RNase H 处理，RNase H 可降解 RNA-DNA 杂交分子中的 RNA 部分。图 14.19 显示，与 U5、内含子和 E2 互补但不与第一个外显子互补的寡聚核苷酸可协助 RNase H 降解复合体。因此，复合体显然包含 U5 和内含子-外显子 2 剪接中间体，且 U5 与第二个内含子的相互作用具有位点特异性，因为用 4-thioU 取代剪接底物的第二个外显子的第二个碱基后，无双分子 RNA 复合体的形成。

为鉴定 U5 或 U6 分子中参与 4-thioU 与剪接底物交联的碱基，Sontheimer 和 Steitz 开发了引物延伸阻断法（primer extension blockage）。研究者以与 snRNP 序列互补的寡核苷酸为引物，对复合体中的 snRNP 进行反转录，当遇到交联位点时，反转录酶即刻停止反转录而产生特定长度的 DNA 片段，其长度相当于引物结合位点与交联位点之间的距离。因此，根据这些信息就可准确地确定发生交联的位点。图 14.20 显示了引物延伸阻断实验结果。图 14.20a 和图 14.20b 表明，不管是以完整的剪接底物为模板还是以第一个外显子为模板，均能证明 U5 中有 2 个相邻的 U 与第一个外显子的最后一个碱基发生了交联。暂略过图 14.20c，图 14.20d 表明，U5 中相邻的 2 个 U 不仅与第一个外显子的最后一个碱基发生交联，而且其中的一个 U 连同其相邻的 C 还与第二个外显子的第一个碱基发生了交联。图 14.20c 显示，U6 中有 4 个碱基与内含子的第二个碱基发生了交联。对 U5 的研究结果表明，U5 snRNP 能与第一个外显子的 3′端及第二个外显子的 5′端结合（见图 14.21 的说明），从而将两个外显子置于利于剪接的位置处。

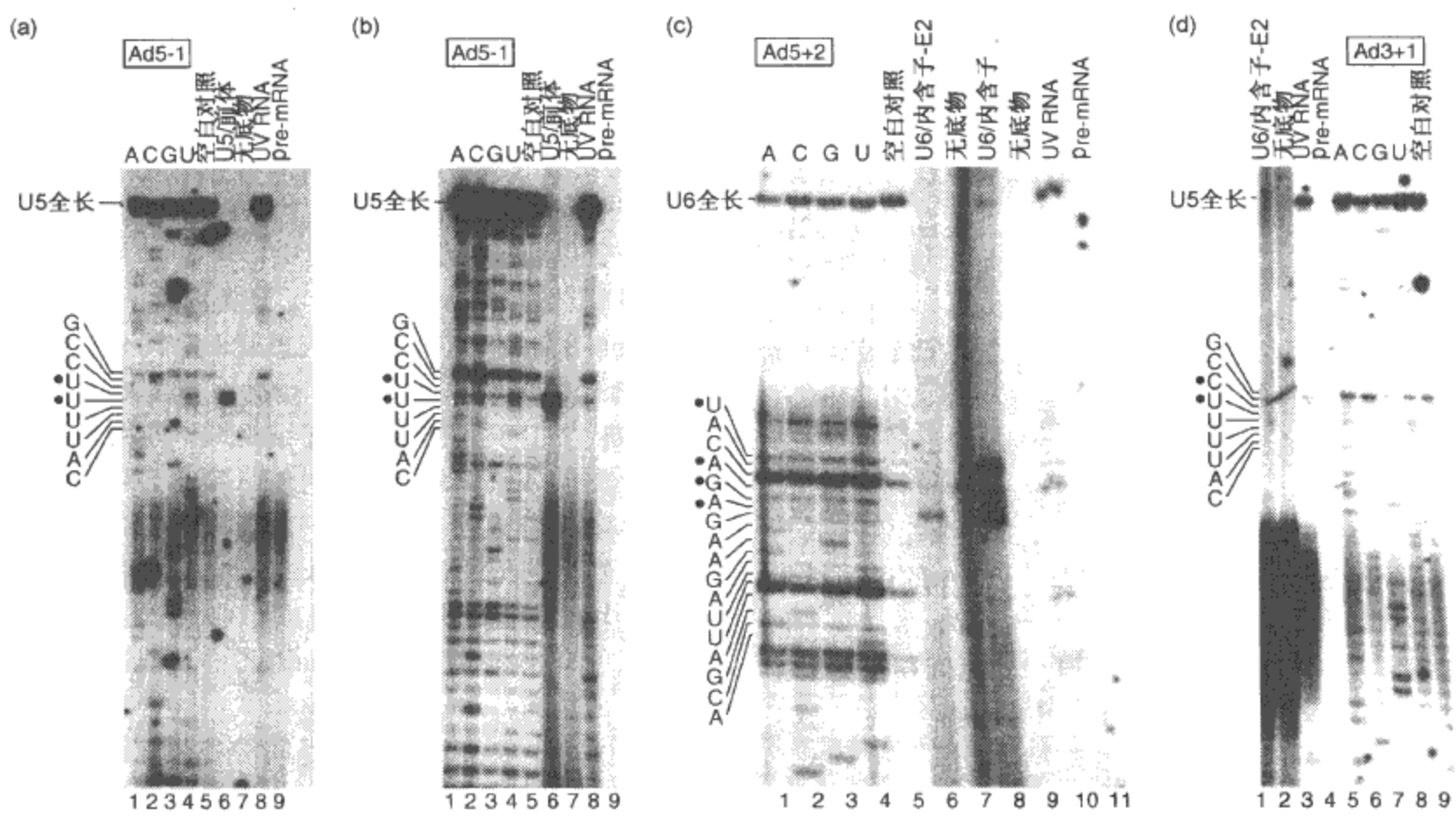


图 14.20　与位于剪接底物不同位点处的 4-thioU 发生交联的 snRNP 碱基鉴定。Sontheimer 和 Steitz 利用引物延伸实验，确定了 U5 和 U6 与 4-thioU 发生交联的碱基：第一个外显子的最后一个碱基［图（a）和（b）中的 Ad5-1］；内含子的第二个碱基［图（c）中的 Ad5＋2］；第二个外显子的第一个碱基［图（d）中的 Ad3＋1］。研究者利用这些 RNA 分子形成交联复合物，通过凝胶电泳分离复合物，添加 U5 或 U6 的特异引物，进行引物延伸实验分析。图（a）、（b）、（c）中的前 4 个泳道以及图（d）中的 5～8 泳道是测序泳道，使用的引物与引物延伸实验相同。标有“空白”的泳道为没有模板的对照测序泳道。图（a）和（b）中的泳道 6，图（c）中的泳道 6、8，图（d）中的泳道 1 为实验泳道。引物延伸实验所使用的复合物分别为：U5/剪接前体复合物［U5/pre8，图（a）］、U5/外显子 1 复合物［U5/E1，图（b）］、U6/内含子-外显子 2 复合物［U6/内含子-E2，图（c）］、U5/内含子-外显子 2 复合物［U5/内含子-E2，图（d）］。其他泳道为对照。“无底物”是反应混合物中没有底物，从对应于有底物时复合物应出现的位置处割下空白胶条；“UV RNA”是无底物提取物中的总 RNA；“pre-mRNA”是未交联的底物。每个图左侧带点的碱基表示 snRNP 上发生交联的位点。（*Source*：Sontheimer，E. J. and J. A. Steitz，The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science* 262（24 Dec 1993）f. 5，p. 1993. Copyright © American Association for the Advancement of Science.）

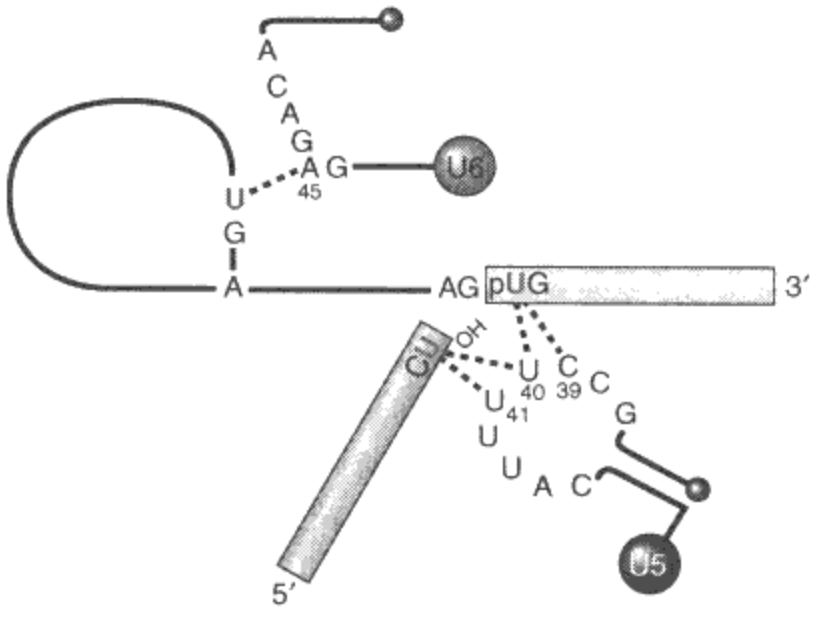


图 14.21　通过 4-thioU 交联作用所揭示的 U5 和 U6 与剪接底物间相互作用的总结。红色粗体 U 代表引入到剪接底物中的 4-thioU，点线表示剪接底物的 4-thioU 与 snRNP 碱基间的交联，外显子 1 为蓝色，外显子 2 为微黄色。紫色小圆点表示 snRNA 5′端的帽子结构。U5 的作用是在第二步剪接时将两个外显子放置在适当位置处。(*Source*: Adapted from Sontheimer, E. J. and J. A. Steitz, The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science* 262: 1995, 1993.)

小结　U5 snRNA 能与一个外显子的最后一个碱基和下一个外显子的第一个碱基发生交联。可能的作用是将这两个外显子排列在一起便于剪接。

U4 snRNP　我们对 U4 的了解主要是关于其与 U6 的结合。U4 snRNA 和 U6 snRNA 的序列提示我们，两者可通过碱基配对而形成两个茎：茎 I 和茎 II，现在已有证据表明 U4 与 U6 之间存在着相互结合作用。事实上，在图 14.16 中已显示出 U4-U6 复合体存在的强烈信号。在剪接过程中，U4 是否具有直接作用？很显然，U4 在剪接过程中并没有发挥直接作用。在剪接开始时，U4 就与 U6 解离，且用温和的方法即可使 U4 与剪接体分离。由此可见，U4 的作用可能是暂时与 U6 结合而使之被封闭，直到 U6 参与到剪接反应中。值得注意的是，U6 序列中与 U4 配对形成茎 I 的碱基也能够与 U2 配对（本章前面讨论过）。U4 解离的重要性在于使 U6 与 U2 发生碱基配对，有助于活性剪接体的形成。

另一个有趣的特点是，U6 snRNA 基因被一个 mRNA 型的内含子所间断。至少有两种酵母存在该内含子，即粟酒裂殖酵母（*Schizosaccharomyces pombe*）和红冬孢酵母（*Rhodosporidium dacryoidum*）。

小结　U4 可与 U6 进行碱基配对，其作用可能是与 U6 结合，直至 U6 需要参与到剪接反应时，U4 才与 U6 解离。至少在两种酵母中，U6 基因被 mRNA 型内含子所间断。

snRNP 参与 mRNA 的剪接　在本章后面我们将了解到，有些类型的内含子能够不依赖于剪接体而进行自我剪接（self-splicing），这些内含子具有自我剪接所需要的所有催化活性。自我剪接的内含子可分为两类，其中 II 类内含子通过套索中间体进行自我剪接，这种套索中间体类似于核 mRNA 剪接过程中的套索中间体。这一特点促使我们做出如下推测：剪接体的 snRNP 可被 II 类内含子形成的相似结构所取代，该结构可将外显子 1 和外显子 2 并列放置在一起，便于剪接。

图 14.22 描述了剪接体剪接模型和 II 类内含子剪接模型。图 14.22a 描述的模型是图 14.21 所示核 mRNA 第二步剪接模型的一种变形。图 14.22b 为 II 类内含子的剪接模型。值得关注的几个特征如下。首先，U5 环取代了 II 类内含子的 ID 结构域，将与之结合的外显子 1 和 2 定位在利于剪接的位置处。ID 结构域的 RNA 序列称为内部引导序列（internal guide sequence），其功能是引导其他 RNA 区域进入合适位置以利于催化剪接。其次，与 5′-剪接位点发生碱基配对的 U6 序列区取代了 II 类内含子的 IC 结构域。再次，U2-U6 配对形成螺旋 I 类似于 II 类内含子的 V 结构域。最后，U2 与分支点配对形成的螺旋取代了 II 类内含子的 VI 结构域。在两种情况下，分支点 A 周围的碱基配对使 A 突出，可能有助于 A 发挥其形成分支的功能。由于 II 类内含子是具有催化功能的 RNA［核酶（ribozyme）］，那么，图 14.22 中两种模型的相似性表明，取代位于剪接活性中心的 II 类内含子元件的 snRNP 也应具有催化剪接反应的功能。

2000 年，Ren-Jang Lin 及同事提供了有关 U6 snRNA 确实参与催化反应的证据。研究者的论据如下。内含子的剪接是通过两次转酯反应完成的（见图 14.4）。转酯反应过程中，在断开一个磷酸二酯键的同时会形成另一个新的磷酸二酯键。例如，在第一步转酯反应中，第一外显子与内含子之间的磷酸二酯键断裂，在分支点 A 与内含子 5′端之间形成一个新的磷酸二酯键，形成套索状中间体。转酯反应中的催化剂具有两种功能：一种是活化亲核试剂（分支点 A 的 2′-OH）；另一种是稳定离开的基团（在第一个外显子末端形成 3′-OH 的氧原子）。金属离子（如镁离子）就是这样的催化剂。II 类内含子的自我剪接确实需要镁离子的参与。

Lin 及同事研究发现，用硫取代 U6 snRNA 中

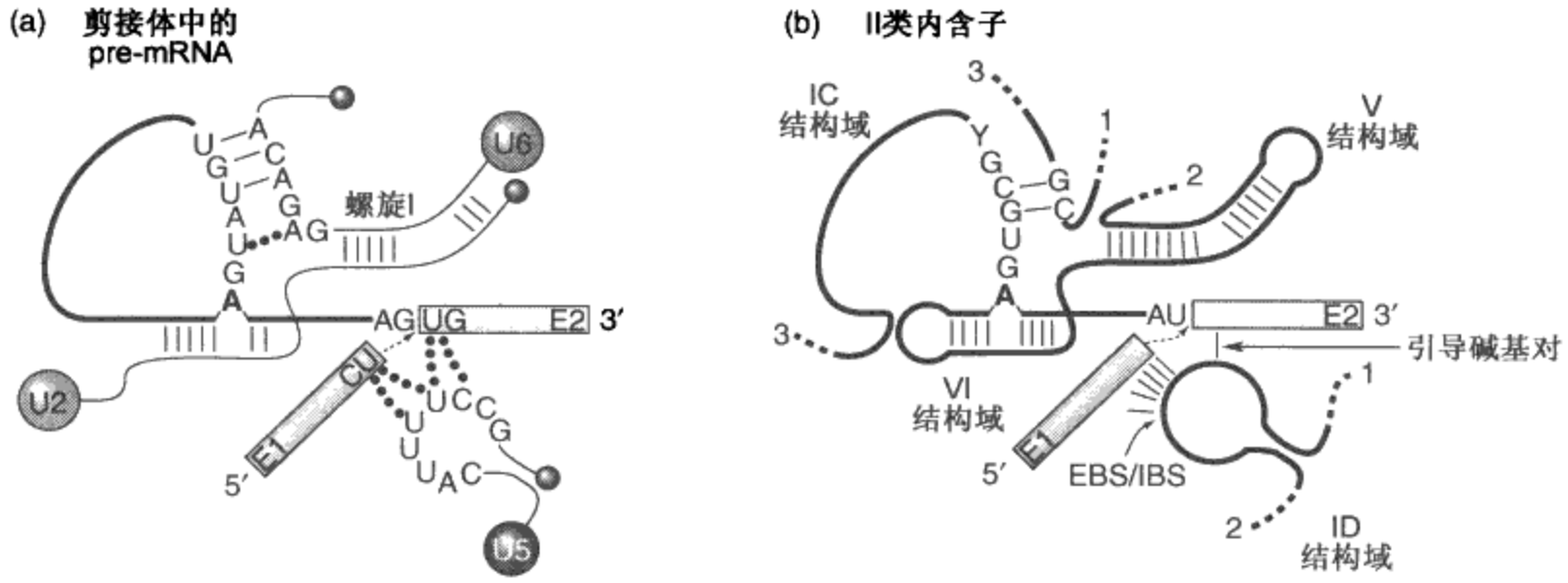


图 14.22　剪接体活性中心模型与 II 类内含子活性中心模型的比较。（a）剪接体。这是图 14.21 模型的一种变形，包含 U2（棕红色），其他颜色表示同图 14.21 一致。粗体 A 为分支点，粗线条代表内含子，虚线箭头表示外显子 1 对内含子-外显子 2 的攻击反应。（b）II 类内含子。内含子的表示方法与图 a 所示剪接体中的相同，只显示了部分内含子序列，其余序列用点线表示，数字指明了各部分间的连接。有色部分为外显子，粗体 A 为分支点，虚线箭头表示外显子 1 对内含子-外显子 2 的攻击反应。（*Sources*：(a) Adapted from Wise, J. A., Guides to the heart of the spliceosome, *Science* 262: 1978, 1993, (b) Adapted from Sontheimer, E. J. and J. A. Steitz, The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science* 262: 1995, 1993.）

的一个氧会完全阻断剪接反应，这种取代产生的剪接阻断效应是可以预知的，因为硫取代氧后会阻碍 U6 与镁离子的结合。如果此关键性镁离子位于催化位点，这就意味着 U6 在催化反应中也发挥直接作用。如果事实确实如此，那么，硫取代氧产生的阻断效应就可被添加的锰离子所逆转，因为锰离子具有同镁离子一样的催化作用。但是，与镁离子不同的是，锰离子可与 RNA 结合，尽管该 RNA 分子中的一个关键氧原子被硫所取代。

Lin 及同事研究发现，U6 snRNA 中因硫取代氧而产生的剪接阻断效应确实可被锰离子所逆转。这表明在剪接体的催化中心，U6 能够与镁离子结合，但这还不足以证明镁离子就位于催化中心内，因为即使不在催化中心，金属离子也能发挥催化作用。

在 2001 年，Saba Valadkhan 和 James Manley 为 RNA 催化假说提供了更多证据。研究者将体外合成的 U2 和 U6 snRNA 片段混合，然后加入含有分支点保守序列的酵母内含子寡核苷酸片段，研究表明这些 RNA 分子可以催化与第一步剪接反应相关的转酯反应。在正常的第一步剪接反应中，由分支位点 A 攻击连接第一个外显子与内含子之间的磷酸二酯键（5′-剪接位点）。在由 U2、U6 和内含子片段所催化的体外反应中，由于没有 5′-剪接位点，分支位点 A 会攻击 U6 自身的磷酸二酯键，形成分支状寡聚核苷酸。图 14.23 显示了在该反应中三类 RNA 分子之间的碱基配对情况、参与催化反应的核苷酸以及产物的可能结构。

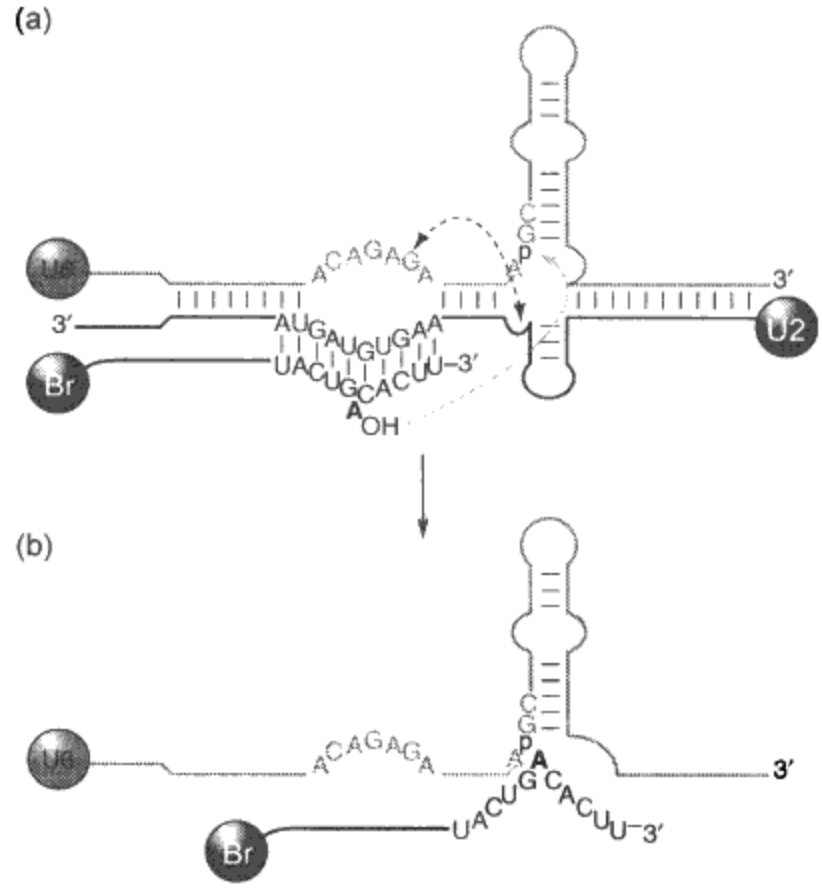


图 14.23　类似于剪接复合体第一步剪接反应的体外反应。（a）在体外，三类 RNA 分子通过碱基配对形成复合体，U6 片段（红色）位于顶部，U2 片段（蓝色）位于中间，带有突出分支点 A（粗体）的分支点片段（Br，黑色）位于底部，灰色箭头指向 A52-G53 两个核苷酸间的磷酸二酯键（黑色），这个磷酸二酯键是分支点 A 的攻击对象。虚线箭头连接的碱基为在紫外光下，U6 和 U2 中能够发生交联的碱基。（b）产物可能的化学结构。

在不同的反应条件下，Valadkhan 和 Manley 将标记的分支位点寡聚核苷酸（Br）添加到含有 U2 和 U6 snRNA 的混合体系中。图 14.24 给出了实验的结果。图 14.24a 显示，在反应进行到 24h 后形成了产物 X，纯化后进行凝胶电泳可显示该产物的存在。X 产物的形成依赖于 U2 和 U6 片段的同时存在，当反应体系被加热到接近 U2-U6 复合体的熔解温度时，X 产物的合成反应被阻断。由此可见，U2 和 U6 片段都是该反应所必需的。图 14.24b 和图 14.24c 显示，在反应进行到 2h 后，X 产物呈线性增加，一直持续到接近 20h，该反应可被添加更多的 U2 和 U6 片段所促进，直至浓度达到 2μmol/L 的饱和水平。

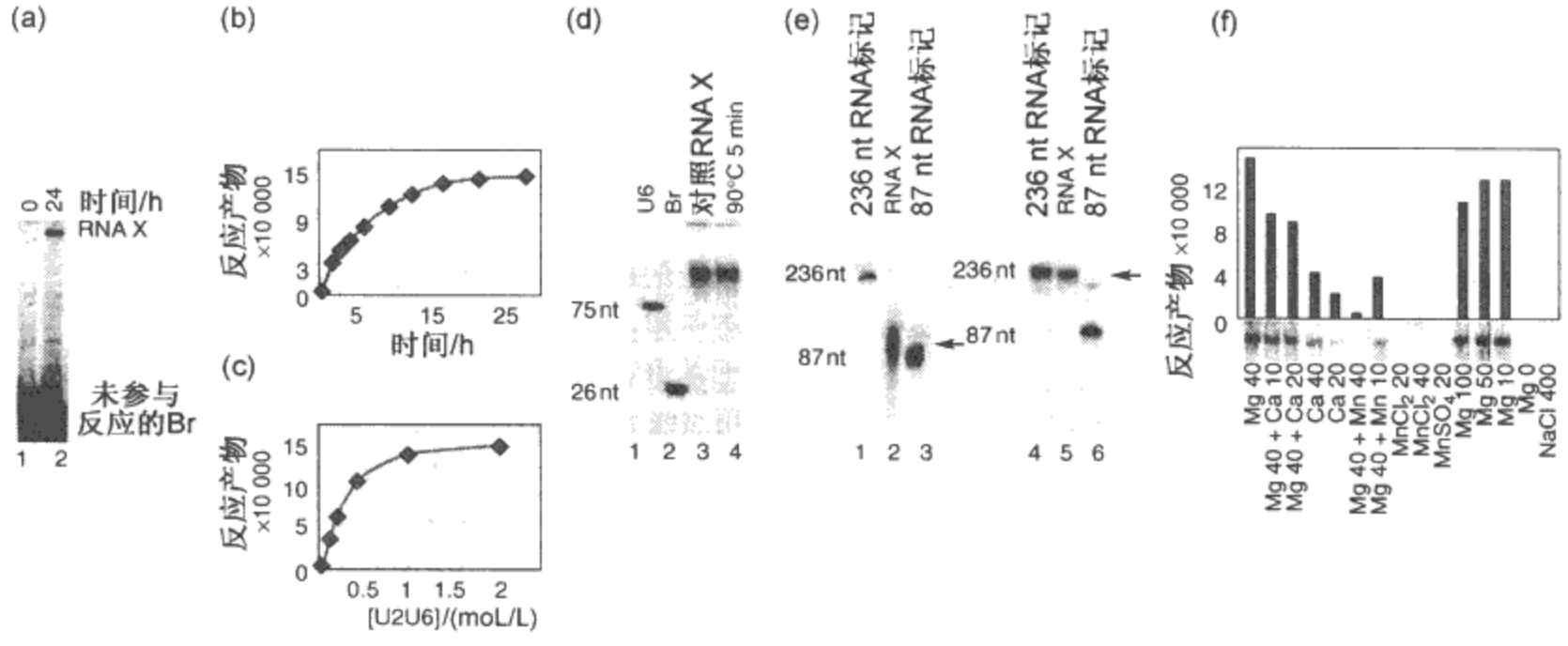


图 14.24　RNA X 产物的形成。（a）通过 SDS-PAGE 方法检测 RNA X。Valadkhan 和 Manley 将体外合成的 U2 snRNA、U6 snRNA 和 Br 片段混合，在 Mg^{2+} 存在的条件下，孵育 0h 或 24h，然后电泳。（b）反应的时间曲线。（c）反应对 U2 和 U6 的依赖性。（d）RNA X 对热熔解反应的抗性。泳道 3 为未被加热的 RNA X 的电泳迁移情况；泳道 4 为 90℃加热 5min 后，RNA X 的电泳迁移情况，两者没有任何差异。泳道 1 和 2 是分别用 U6 和 Br 片段所做的对照实验。（e）未被加热的 RNA X 产物在 8%（泳道 1～3）及 16%（泳道 4～6）的聚丙烯酰胺凝胶上的泳动行为异常。（f）不同浓度的各种二价金属离子对反应的影响。金属离子的浓度标注在各泳道的下方，各反应的产物量标注在每个泳道的上方。（*Source*：Reprinted with permission from *Nature* 413：from Valadkhan and Manley fig. 2，p. 702. © 2001 Macmillan Magazines Limited.）

相同的系列实验结果也表明，RNA X 产物可能含有分支点核苷酸。图 14.24d 显示，RNA X 不可能由 U2 和 U6 通过强碱基配对形成的，因为 RNA X 能够抵抗 90℃的高温处理，所以 RNA X 的形成显然涉及 RNA 分子间的共价结合，而不只是 RNA 分子间的碱基配对。图 14.24e 表明，RNA X 具有异常的电泳行为。在 8%聚丙烯酰胺凝胶上，RNA X 迁移的位置恰好位于 87nt 分子标记的上方（左侧），在 16%聚丙烯酰胺凝胶上，RNA X 迁移的位置恰好位于 236nt 分子标记的下方（右侧），显示出分支状 RNA 分子的电泳特征（见图 14.5）。图 14.24f 表明，RNA X 产物的形成依赖于 Mg^{2+} 的存在，Ca^{2+} 能够取代 Mg^{2+} 的作用，但效率不高，而 Mn^{2+} 对该反应根本无促进效应。

接着，Valadkhan 和 Manley 标记 Br 及 U2 和 U6 片段的 5′端和 3′端，反应结束后发现，RNA X 出现来自 U6 和 Br 的末端标记信号，但没有 U2 的标记信号。所以，RNA X 的组分包含 U6 和 Br，但没有 U2。同时，U6 和 Br 不仅通过碱基配对连接在一起，而且这两个 RNA 分子还可能发生共价连接。Valadkhan 和 Manley 的研究还表明，封闭 Br 和 U6 的 5′端（分别进行脱磷酸和引入环磷酸盐）并不能抑制 RNA X 的形成。由此可见，这两个 RNA 分子的 5′端没有参与分子间的连接，连接一定发生在 RNA 分子的内部，从而产生“X”形产物。

最后，Valadkhan 和 Manley 定位了这两个 RNA 分子中与 Br 的分支点 A，以及与 U6 固定序列 AGC 中 A53 和 G54 之间的磷酸发生连接的位点（见图 14.23）。为定位连接位点，他们采用相同的引物延伸分析法，即在分析 4-thioU 与 U5 和 U6 及剪接底物的交联位点时采用的定位方法（图 14.20），同时对末端标记的 RNA X 进行化学切割处理，以鉴定因 RNA-RNA 相互作用而阻止切割的核苷酸。

这些实验的结果表明，在没有蛋白质协助下，

Mg^{2+}、U2、U6 和 Br 能够催化类似于第一步剪接反应。当然，这个反应与第一步剪接反应并不完全相同，因为底物中没有可供分支点 A 攻击的 5′-剪接位点。然而，被攻击位点存在于 U6 上情况并不是没有先例的：在体外，有时会发生异常剪接，被攻击位恰好位于 U6 的碳骨架上。在酵母 U6 基因上确实发现有一个内含子插入到邻近保守序列 AGC 的位点处，这种插入可能是分支位点 A 对 U6 而不是 5′-剪接位点进行异常攻击的结果。

综上所述，这些研究结果强烈表明，剪接体催化中心涉及 Mg^{2+}、能够进行碱基配对的 U2 snRNA、U6 snRNA 和内含子分支点序列。在体内也可能需要蛋白质的参与，但在体外，至少在这些实验条件下，催化中心显然不需要蛋白质的参与。

小结 即将进行第二步剪接反应的剪接体（底物、U2、U5 和 U6）的结构特征可以从处于相同剪接阶段的 II 类内含子的结构特点推知。因而，II 类内含子中位于催化活性中心的元件似乎可以被剪接体的 snRNP 所取代，这些元件可能具有剪接体的催化活性。剪接体催化中心显然包括 Mg^{2+} 和通过 U2 snRNA、U6 snRNA 和内含子分支点序列间碱基配对形成的复合体。这三类无蛋白质组分的 RNA 片段可催化与第一步剪接有关的反应。

剪接体的组装及功能

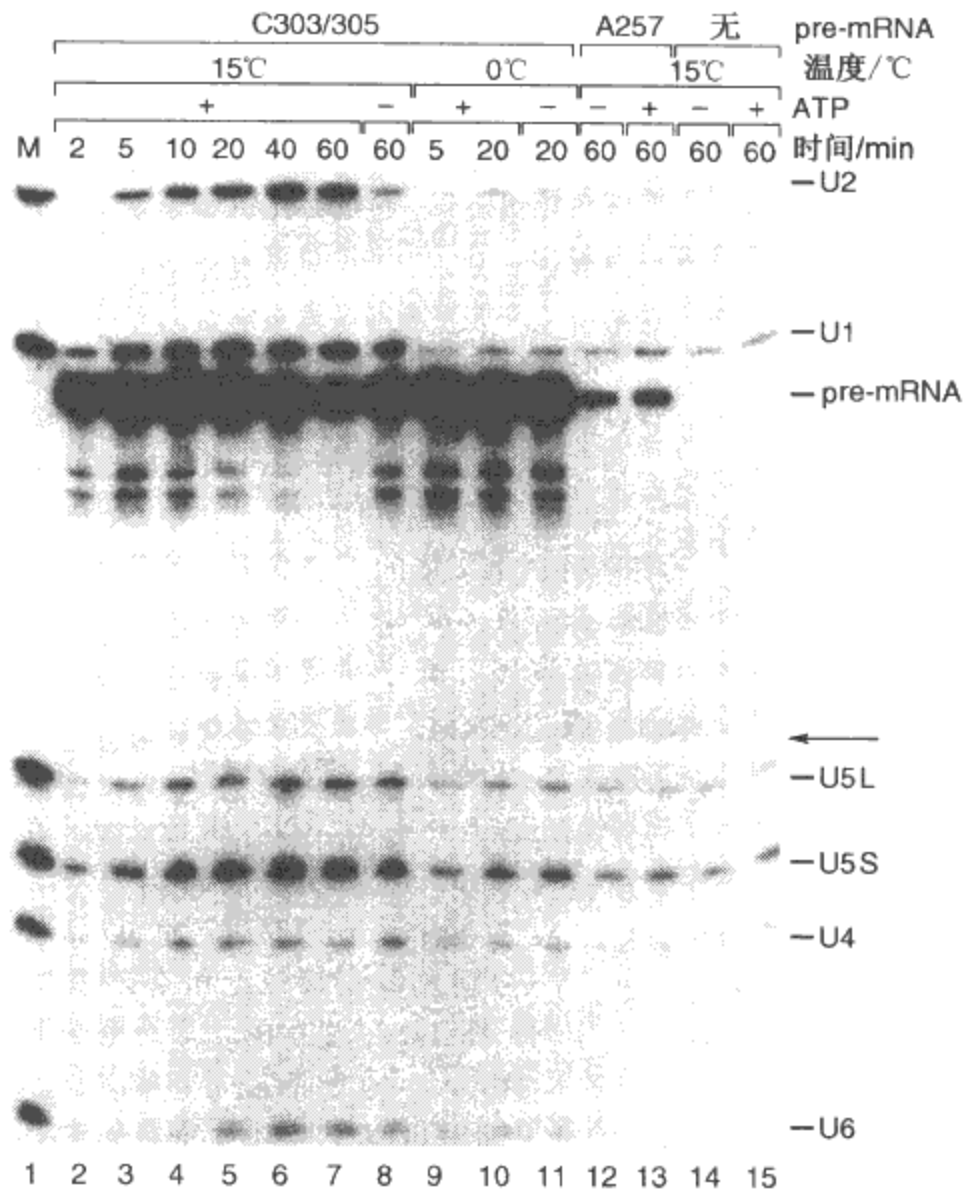
剪接体由多种组分逐级组装而成，包括蛋白质和 RNA。剪接体组装的部分次序已经明确。我们将剪接体的组装、功能行使、解体称为**剪接体循环**（spliceosome cycle）。本节主要讨论剪接体循环，并了解细胞通过对剪接体组装过程的控制，调节剪接反应的质量和数量，实现调控基因表达的目的。

剪接体循环 在多个研究小组首次分离出剪接体之时，出人意料的是，研究者并没有发现 U1 snRNP 的存在，而 U1 确实能够通过 RNA-RNA 间的碱基互补配对而与 5′-剪接位点结合，是起始剪接所必需的。U1 的确是剪接体的组成部分，但可能由于当时纯化剪接体的方法过于苛刻，导致 U1 发生解离。Stephanie Ruby 和 John Abelson 于 1988 年发现，U1 是最先与剪接前体结合的 snRNP。他们设计了一个十分巧妙的实验方案来研究剪接体的组装过程。首先通过生物素-抗生物素蛋白连接（biotin-avidin linkage）将锚定 RNA 结合在琼脂糖珠上，然后使酵母 pre-mRNA与锚定 RNA 进行杂交，从而将酵母 pre-mRNA 固定在琼脂糖珠上。在不同时间段内添加酵母细胞核提取物，洗除未结合的物质，然后洗脱 RNA，进行电泳、印迹，然后用针对各剪接体 snRNA 的探针进行杂交，以检测 snRNA 的种类。

图 14.25 显示的实验结果表明，U1 是最先与剪接底物结合的 snRNP。通过对比图 14.25a 中的泳道 2 和 5 可以看出，在 2min 时间点，U1 是唯一与 pre-mRNA 结合的 snRNP。图 14.25a 同时也表明，除 U1 外，其他所有 snRNP 与剪接底物间的最佳结合都需要 ATP。图 14.25b 是剪接底物结合各类 snRNA 的数量与时间的函数曲线图，很明显，U1 是第一个组装到剪接体上的 snRNP。

为深入探讨剪接体的组装次序，研究者将细胞核提取物，与 U1 snRNA、U2 snRNA 关键序列互补的 DNA 寡聚核苷酸序列及 RNase H 混合孵育，从而使 U1 或 U2 失活。然后进行与前一个实验方案相同的剪接体组装实验。因为 RNase H 能降解 RNA-DNA 杂交分子中的 RNA 组分，所以，与 DNA 寡聚物杂交的 snRNA 分子部分序列会被 RNase H 降解。在降解反应完成后，就可以筛选出与 pre-mRNA（分别为 5′-剪接位点和分支点）杂交的 snRNA 序列。图 14.26 显示的结果有两个要点：①如预期的那样，U1 的失活阻止了 U1 与 pre-mRNA的结合，同时也阻止了其他所有 snRNP 与 pre-mRNA 的结合（对比泳道 2 和 4 的结果）；②与预期一致，U2 的失活阻止了 U2 与 pre-mRNA 的结合，同时也能阻止了 U5 与 pre-mRNA 的结合，但不能阻止 U1 与 pre-mRNA 的结合（对比泳道 2 和 6 的结果）。综合以上结果可以看出，U1 首先结合，然后 U2 在 ATP 协助下完成结合，随后是其他 snRNP 在剪接体上的结合。

U6 一旦与 U4 分离，U6 便取代 U1 与 5′-剪接位点结合，相关内容将在本章后面再做讨论。从其他实验结果中我们能够得知，当 U1 被置换，与 U4 的协同脱离剪接体后，能够启动只包含 U2、U5 和 U6 的剪接体活性。事实上，U6 取代 U1 可能是激活剪接体并启动剪接反应的事件，1999 年。Stephanie Ruby 和 John Abelson 通过实验表明，当改变 5′-剪接位点的碱基从而增强其与 U1 的碱基配对能力时，剪接复合体的激活便被阻断。可能的原因是，因配对能力的增强，U6 很难与 U1 竞争 5′-剪接位点，结果是 U1 与 U4 的分离被阻断，剪接反应也被抑制。相反，保持 U1 与 5′-剪接位点的碱基配对能力不变，而增强 U6 与 5′-剪接位点的碱基配对能力后，剪接体易被激活（释放 U1 和 U4），导致更多剪接事件的发生。Staley 和 Guthrie 的进一步研究表明，U6 与 U1 在 5′-剪接位点发生的置换，需要 U5 snRNP 的蛋白组分 Prp28 和 ATP 的参与。



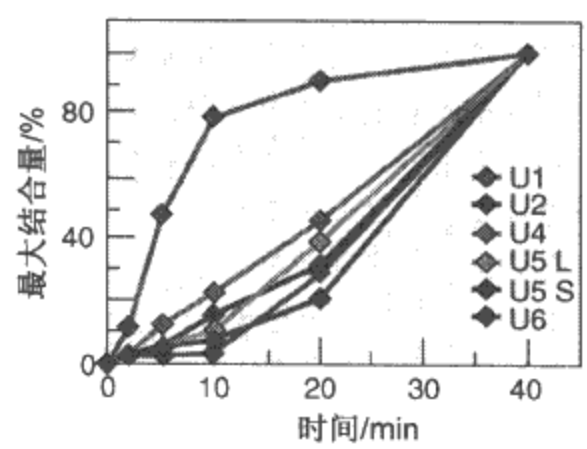


图 14.25　剪接体 snRNP 与 pre-mRNA 结合的动力学。(a) Northern 杂交。Ruby 和 Abelson 通过生物素-抗生物素蛋白之间的连接将锚定 RNA 固定在琼脂糖珠上，然后让酵母肌动蛋白基因的 pre-mRNA 与锚定 RNA 杂交，从而将 pre-mRNA 吸附在琼脂糖珠上。然后将吸附有 pre-mRNA 的小珠与酵母核提取物混合，在有或无 ATP 存在时，分别于 15℃和 0℃温度下孵育 2～60min，如图上方所示。在 3′-剪接位点（C303/305）突变的 pre-mRNA 可组装成剪接体，而保守分支点（A257）突变的 pre-mRNA 不能组装成剪接体。标注“无”的泳道是指无 pre-mRNA，只有锚定 RNA。孵育实验完成后，洗除未结合的成分，从复合物中分离 RNA，然后电泳、吸印，利用 U1、U2、U4、U5 和 U6 特异探针进行 Northern 杂交。鉴定出 U5 的两种形式（U5 L 和 U5 S)。泳道 15：无 pre-mRNA，显示绝大多数 snRNA 的背景结合，作为其他实验组的对照。U1 首先结合，然后其他 snRNP 才开始结合，在分支点（A257）突变的 pre-mRNA 上无任何 snRNP 的显著结合，所有 snRNP 包括 U1 和 U4，在反应进行 60min 后仍然保持结合状态。(b) 各 snRNA 与复合体结合的量与时间的函数关系图。很显然，U1（红色）首先结合，随后才是其他 snRNP 的结合。(*Source*: Ruby, S. W. and J. Abelson, An early hierarchic role of U1 small nuclear ribonucleoprotein in spliceosome assembly. *Science* 242 (18 Nov 1988) f. 6a, p. 1032. Copyright © American Association for the Advancement of Science.)

图 14.27 举例说明了酵母剪接体的循环过程。首先，剪接底物、U1，可能还有其他一些物质组成**定向复合体**（commitment complex，CC)；如其名称所暗示的那样，确保复合体将内含子切除。其次，在 ATP 协助下，U2 加入到定向复合体上形成 A 复合体。再次，U4-U6 和 U5 加入形成 B1 复合体，然后 U4 与 U6 解离，导致：①U6 取代 U1 与 5′-剪接位点结合，激活剪接体，这是一个 ATP 依赖的反应；②U1 与 U4 脱离剪接体；③U6 与 U2 发生碱基配对，被激活的剪接体称为 B2 复合体。在 ATP 供能的条件下完成第一步剪接反应，使两个外显子分离并形成具套索状剪接中间体，外显子和剪接中间体均结合在 C1 复合体上；同样在 ATP 供能条件下发生第二步剪接反应，套索状内含子被切离释放，两个外显子连在一起，这步反应都在 C2 复合体中进行。下一步即是剪接成熟 mRNA 的释放，此时结合有套索状内含子的复合体称为 I 复合体；最后，I 复合体解体，释放出 snRNAP 组分，用于另一个剪接

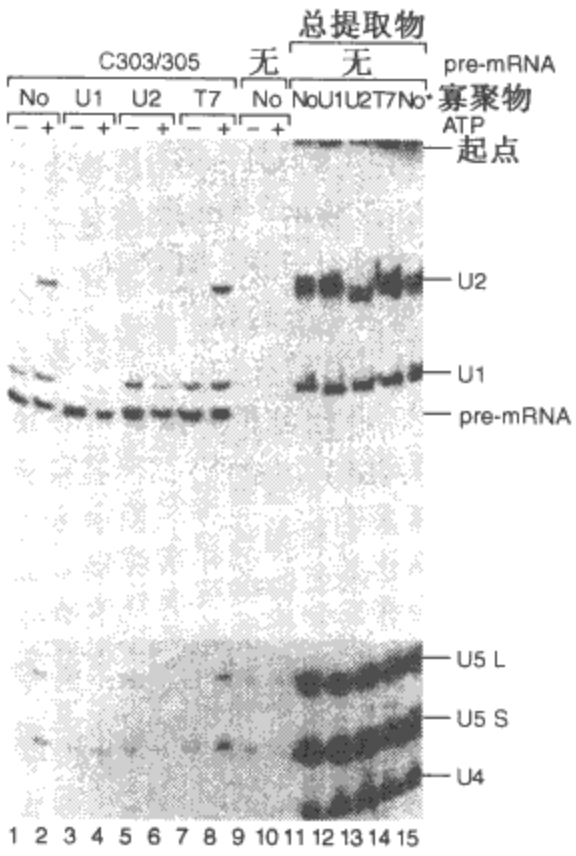


图 14.26　U1 或 U2 失活对剪接体组装的影响。Ruby 和 Abelson 将 U1 或 U2 与 RNase H、能与 U1 snRNA 或 U2 snRNA 的关键区互补的 DNA 寡聚核苷酸保温孵育，使 U1 或 U2 失活。泳道 11～15 是经 RNase H 处理的提取物中标记 snRNA 的电泳结果。不加入寡聚核苷酸（No）、U1 反义寡核苷酸（U1）、U2 反义寡核苷酸（U2）、T7 反义寡核苷酸（T7）。后一个 T7 泳道为第二阴性对照。RNase H 和 anti-U1 组合处理导致 U1 snRNA 由全长变为截短的形式，其泳动速度比全长亲本分子快一些；RNase H 和 anti-U2 组合处理导致 U2 snRNA 几乎全被降解，只出现少量截短的 U2 snRNA。泳道 1～10 显示了剪接体组装实验的结果。实验过程同图 14.25。实验条件在图上方标注：C303/305 pre-mRNA 或无 pre-mRNA；RNase H 处理的提取物中：无寡核苷酸、加入 U1 反义寡核苷酸、加入 U2 反义寡核苷酸、加入 T7 反义寡核苷酸；加入或不加入 ATP。失活 U1 会阻止 U1、U2 和 U5 的组装；失活 U2 会阻止 U2 和 U5 的组装。（*Source*：Ruby, s. W. and J. Abelson, An early hierarchic role of U1 small nuclear ribonucleoprotein in spliceosome assembly. *Science* 242 (18 Nov 1988) f. 7, p. 1032. Copyright © American Association for the Advancement of Science.）

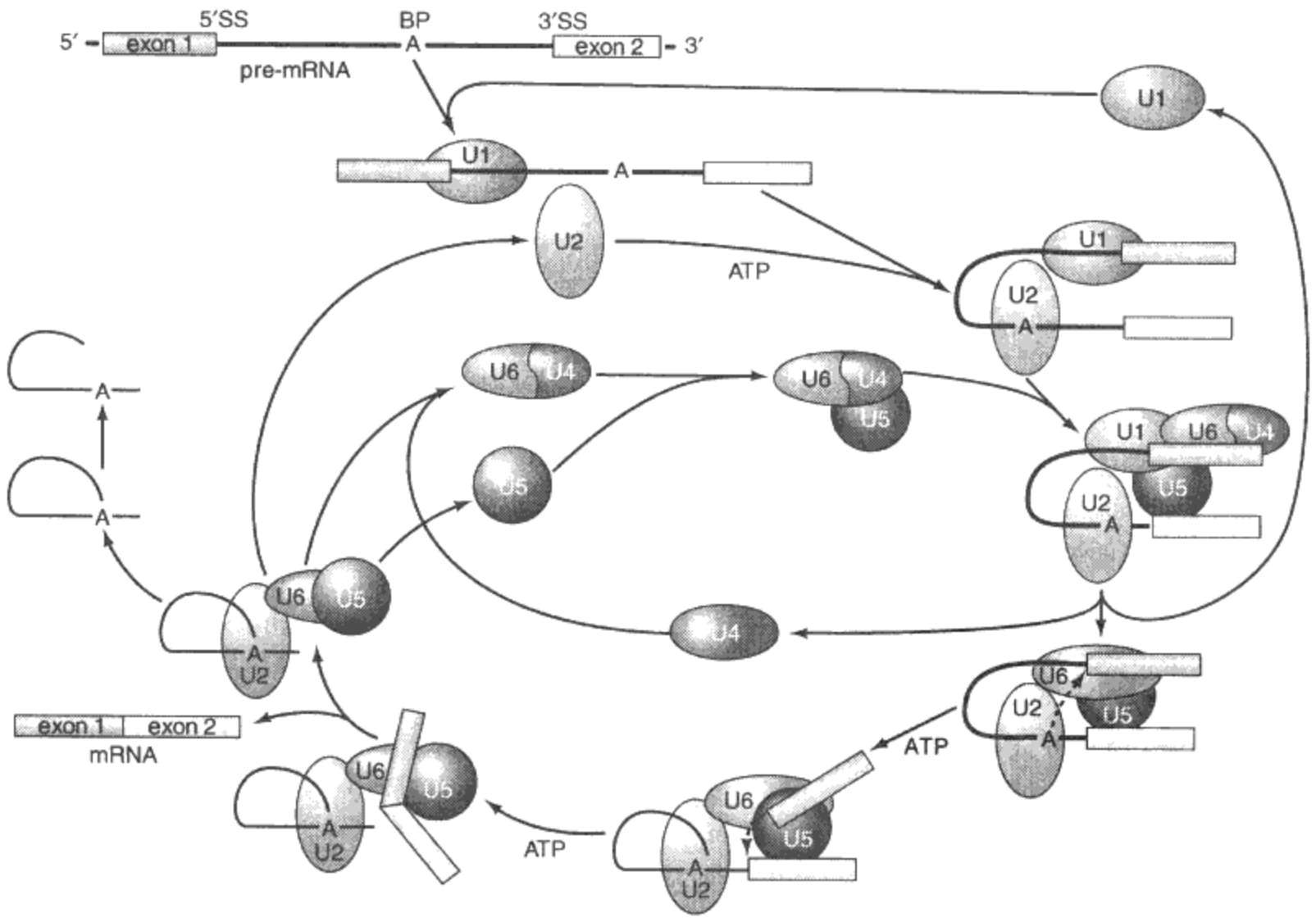


图 14.27　剪接复合体循环。循环过程中的各个事件已在文中阐述。（*Source*：Adapted from Sharp, P. A. Split genes and RNA splicing. *Cell* 77：811，1994.）

体的组装。套索状中间体被切开形成线性分子，很快被降解。

小结　剪接体循环包括剪接体的组装、剪接活性和解体。组装起始于 U1 与剪接底物的结合形成定向复合体。U2 是第二个加入到复合体中的 snRNA，随后是其他 sn RNA 的结合。其中 U2 的

结合需要 ATP 供能。当 U6 与 U4 分离后，U6 便取代 U1 与 5′-剪接位点结合，这个 ATP 依赖的过程激活剪接体，并释放 U1 与 U4。

snRNP 结构 所有的 snRNP 均含有一套相同的蛋白质——7 个 **Sm 蛋白**（Sm protein），这些蛋白质均能被一种来自系统性自身免疫疾病（systemic autoimmune disease)(如系统性红斑狼疮病）患者体内的抗体所识别。snRNA 结合 Sm 蛋白的位点称为 **Sm 位点**，其保守序列是 AAUUUGUGG。除 Sm 蛋白外，每一类 snRNP 还有一套自身特异的蛋白质。例如，U1 snRNP 含有 70K、A、C3 种蛋白质，分子质量分别为 52kDa、31kDa、17.5kDa。

Holser Stark 及同事应用单离子冷冻电子显微镜技术获得了分辨率为 10Å 的 U1 snRNP 晶体结构图像。该结构（图 14.28）显示，Sm 蛋白为环状结构，中央有一个扁平的漏斗状孔洞。两个最大的 U1 特异蛋白 70K 和 A 附着在 Sm 的漏斗处，并与 U1 snRNA 的茎环结构结合。对缺失 70K 或 A 蛋白的 U1 snRNP 进行负染色，通过电子显微镜可鉴定这些突出物的特征，从而判断突出物所对应的蛋白质。

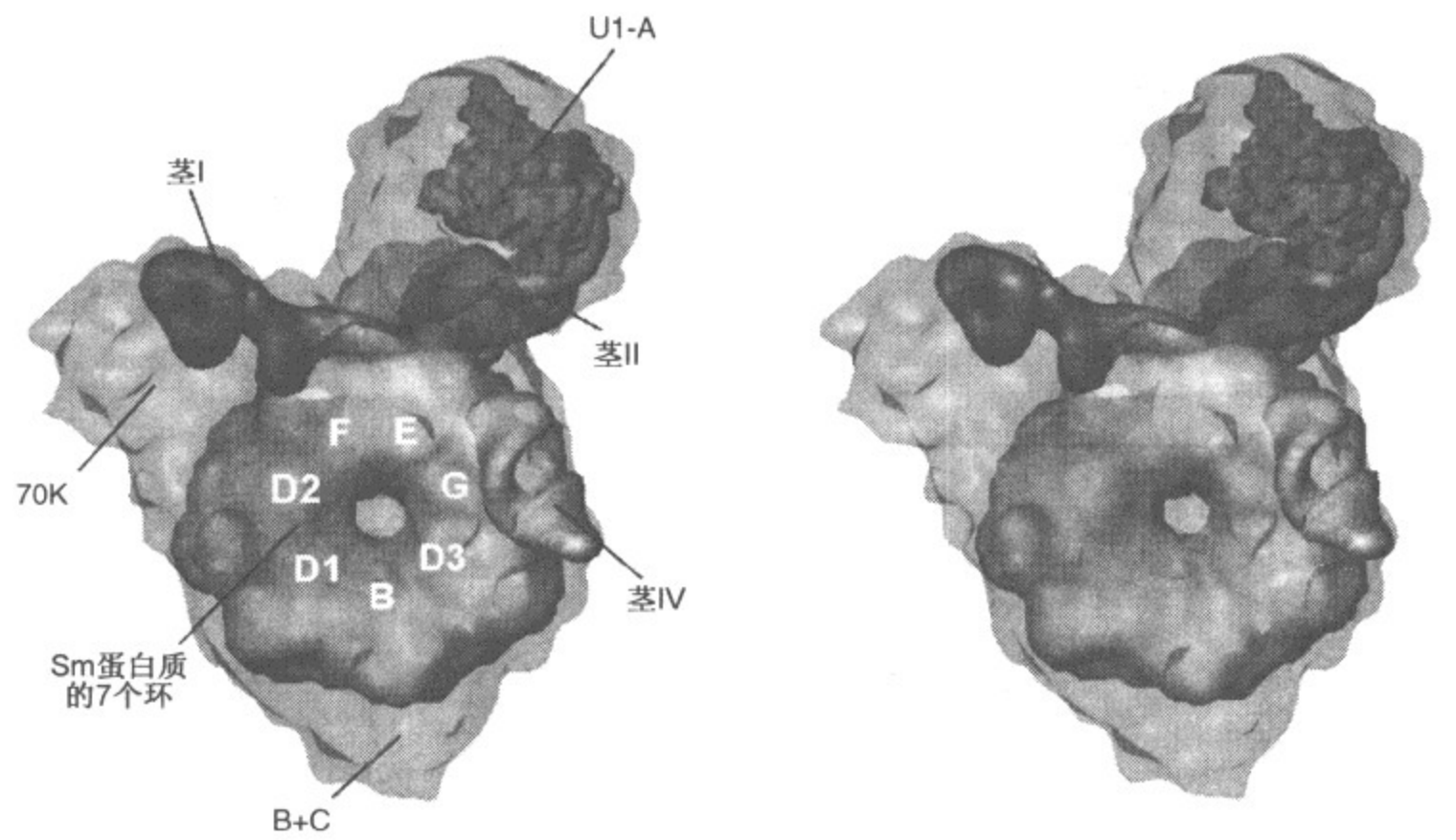


图 14.28 U1 snRNP 的结构。Holser Stark 及同事应用单离子冷冻电子显微镜技术获得了 U1 snRNP 的立体结构图像。主要突出物包括 U1-A 和 70K 蛋白从环形结构 Sm 的中部向外突。(*Source*: Reprinted with permission from *Nature* 409: from Stark et al., fig. 2, p. 540 © 2001 Macmillan Magazines Limited.)

RNA 的 Sm 结合位点处于单链状态，可以穿过环形结构的中央孔洞。事实上，先前通过 X 射线衍射晶体学方法分析 Sm 蛋白的组分时，就曾推测环形结构的孔洞中排列有碱性氨基酸侧链，孔洞的碱性特征有助于 Sm 与 RNA 结合。

小结 参与剪接的 5 类 snRNP 均含有相同的 7 个 Sm 蛋白和一套自身特异蛋白质。U1 snRNP 的结构显示，Sm 蛋白为环形结构，可以结合其他蛋白质。

低丰度剪接复合体（minor spliceosome） 20 世纪 90 年代中期，在后生动物（metazoan）（具有不同器官的动物）中发现了一类稀有的内含子变体，其 5′-剪接位点序列和分支点序列是高度保守的，这与大多数内含子共有序列保守性相对较弱的特点有很大的不同。该发现引发的问题是：如果已知的 snRNA 特别是 U1 和 U2 的序列不能与内含子序列匹配，那么含有这类变体内含子基因的转录物能否被剪接？其实，后生动物细胞含有参与这类内含子剪接的低丰度 snRNA：**U11、U12、U4atac、U6 atac**，在功能上分别与 U1、U2、U4、U6 相似。

这类可变剪接系统的存在可以用于研究 snRNA 与 pre-mRNA 关键位点碱基配对的重要性。事实上，变体 U11 snRNA 可与 5′-剪接位点配对；U12 snRNA 可与分支位点配对；而且，如同 U4 可与 U6 碱基配对一样，U4 atac 与 U6 atac 也能进行碱基配对。

与这类低丰度 snRNA 相关的蛋白质如何与 snRNA 结合形成 snRNP？首先值得关注的是，细胞

内除了存在游离的 U11 snRNP 和 U12 snRNP 外，二者还能以单体的形式结合形成 U11/U12 snRNP。与 U11 snRNP 或 U12 snRNP 相关中的结合蛋白，有些是特异的，有些与高丰度 snRNP 相同，如 Sm 蛋白就存在于所有高丰度 snRNP 中。

小结 一类少数变体内含子的 5′-剪接位点序列和分支点序列是高度保守的。U11、U12、U4 atac 和 U6 atac 参与这类内含子的剪接。

定向、剪接位点的选择以及可变剪接

snRNP 自身没有足够的特异性和亲和力，使其专一而紧密地结合在外显子-内含子边界处，从而使转录物的外显子与内含子区分开来。因此，snRNP 与剪接底物的结合需要额外剪接因子的帮助。而且，需要某些剪接因子能够横跨在内含子和外显子之间，从而界定这些 RNA 元件以利于剪接。在本节，我们主要学习这些剪接因子以及它们如何确保在特定位点处发生剪接，然后再了解另一些因子如何能够从一个剪接位点转移到另一个剪接位点。

3′-剪接位点的选择 在第二步剪接反应中，外显子 1 的 3′-OH 基团攻击内含子末端 AG 与外显子 2 第一个核苷酸之间的磷酸二酯键，AG 位于分支点下游 18～40nt 处最为理想。较靠近分支点的 AG 通常会被剪接机构漏过。那么，如何确定用于剪接的正确 AG 位点？Robin Reed 及同事发现，**剪接因子**（splicing factor）**Slu7** 是选择正确 AG 位点所必需的。如果没有 Slu7 的参与，剪接体可能就会在错误的 AG 位点处对剪接底物进行剪接。

Katrin Chua 和 Robin Reed 利用交联有抗 Slu7 抗血清的琼脂糖珠，免疫剥离 HeLa 细胞提取物中的 Slu7 蛋白，然后使提取物与琼脂糖珠分离，从而获得无 Slu7 的 HeLa 细胞提取物。作为对照，研究者用交联有无抗 Slu7 抗体的免疫前血清的琼脂糖珠处理 HeLa 细胞提取物，其中的 Slu7 蛋白并没有缺失。然后，以标记的模式 pre-mRNA 为剪接底物，检测这些处理后的 HeLa 细胞提取物的剪接能力，模式 pre-mRNA 是腺病毒主要晚期基因转录物的修饰产物，只包含一个 AG 信号且位于分支位点下游 23nt 处（图 14.29）。模式剪接底物与提取物共同孵育后，Chua 和 Reed 电泳反应产物以检测提取物对底物的剪接情况。

图 14.29a 显示了“天然”底物剪接的实验结果。未缺失 Slu7 的提取物能够完成两步剪接反应，产生成熟的 mRNA、内含子以及少量未被剪接的外显子。此外，缺失人 Slu7 的提取物（△hSlu7）只

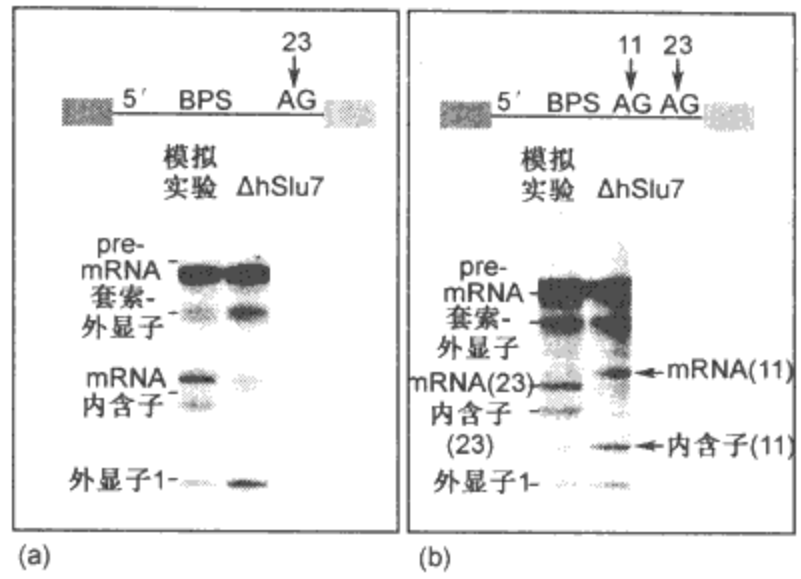


图 14.29 Slu7 蛋白是在 3′-剪接位点正确的 AG 处发生剪接所必需的。Chua 和 Reed 检测了 HeLa细胞提取物选择 3′-剪接位点 AG 的能力。一组实验所用的提取物中含有 Slu7 蛋白（模拟），另一组实验所用的提取物中不含有 Slu7 蛋白（△hSlu7，利用抗 Slu7 抗血清免疫剥离 Slu7 蛋白）。标记的剪接底物为腺病毒主晚期基因的 pre-mRNA，包括前两个外显子和第一个内含子。电泳反应产物，放射自显影检测。剪接底物和剪接产物的电泳位置在图左侧标出。（a）剪接底物所含有的 AG 信号位于分支位序列（BPS）下游 23nt 处，本应在正常 AG 处发生的剪接因提取物中缺失 Slu7 而被抑制；（b）剪接底物包含两个 AG 序列，分别位于分支点下游 11nt 处和 23nt 处，在提取物无 Slu7 时，剪接发生分支点下游 11nt 的 AG 处，当提取物中含有 Slu7 时，剪接反应很少在此位点发生。(*Source*: (*photos*) Chua, K., and Reed, R. The RNA splicing factor hSlu7 is required for correct 3′ splice-site choice. *Nature* 402 (11 Nov 1999) f. 1, p. 208. © Macmillan Magazines Ltd.)

能完成第一步剪接反应，第二步剪接反应被阻断，几乎无成熟 mRNA 或内含子的产生，但产生了大量的外显子 1 和套索状内含子-外显子 2。这就意味着 Slu7 是识别 3′-剪接位点正常 AG 信号所必需的。

接下来，Chua 和 Reed 提出这样一个问题，即如果在分支点序列的下游 11nt 处插入一个额外的 AGG 信号，会发生怎样的剪接结果？图 14.29b 显示，未缺失 Slu7 的细胞提取物主要在分支位点下游 23nt 的正常 AG 剪接信号处剪接 pre-mRNA，产生成熟的 mRNA，极少在靠近分支位点非正常 AG 处发生剪接。相反，缺失 Slu7 的细胞提取物主要在非正常 AG 处剪接 pre-mRNA，极少在正常 AG 处发生剪接。在进一步实验研究中发现，当剪接底物的两个 AG 分别位于分支点下游 11nt 和 18nt 处或 9nt

和 23nt 处时，缺失 Slu7 的细胞提取物具有相同的异常剪接行为，既可以在位于正常剪接信号上游的 AG 处也可以在位于正常剪接信号下游的 AG 处发生非正常剪接，而正常的剪接反应却很少发生（在所有的剪接情况中，作为异常剪接的靶点，错误的 AG 剪接信号与分支位点的间隔必须在 30nt 之内）。由此可见，Slu7 不仅是识别正确剪接位点 AG 所必需的，而且在正确剪接位点发生的剪接反应也会因 Slu7 的缺失而受到特异性抑制。

是什么原因导致对异常 3′-剪接位点的选择？Chua 和 Reed 在剪接反应的不同阶段纯化剪接体，组分分析发现，缺失 Slu7 的细胞提取物形成的剪接体无外显子 1，至少在一定的条件下如此。因此，研究者推论，外显子 1 只是松散地结合在缺失 Slu7 提取物形成的剪接体上，这种松散结合的外显子 1 不能与正确的 AG 位点进行剪接，可能是由于该 AG 以某种方式隐藏于剪接体的活性位点内，使外显子 1 不能接近于正确的 AG 位点，因而这个松散结合的外显子 1 只会与邻近的 AG 发生剪接。

另一个参与识别 3′-剪接位点的剪接因子是 U2AF（U2 相关因子，U2-associated factor），该蛋白质由 U2AF35（35kDa）和 U2AF65（65kDa）两个亚基组成。大亚基结合在分支点下游多聚嘧啶碱基区，小亚基与 3′-剪接位点的 AG 结合。

小结 剪接因子 Slu7 是选择正确 3′-剪接位点所必需的。当 Slu7 缺失时，在正确 3′-剪接位点 AG 处的剪接受到特异性抑制，而发生在距分支点 30nt 之内错误 AG 处的异常剪接则被激活。对 3′-剪接位点的识别还需要 U2AF 的参与，65kDa 的 U2AF 亚基结合在分支点下游多聚嘧啶碱基区，35kDa 的 U2AF 亚基与 3′-剪接位点的 AG 结合。

定向 其他一些剪接因子在定向中发挥着关键性作用。1993 年，Xiang-Dong Fu 以人类 β-球蛋白的 pre-mRNA 为剪接底物，发现单体剪接因子 SC35 能够促使定向复合体的形成。Fu 的定向分析实验如下：先将标记的剪接底物和纯化的 SC35 进行预孵育，然后加入细胞核提取物，孵育 2h 以保证剪接反应的发生，最后电泳标记的 RNA，观察是否出现剪接的 mRNA。

图 14.30 显示了实验结果。首先，向反应体系中加入 40 倍过量的具有 5′-剪接位点的非标记 RNA 分子后，能够阻止标记的 β-球蛋白 pre-mRNA 的剪接，这可能是由于两类 RNA 分子对某些剪接因子竞争的结果（对比泳道 1 和 4），但具有 3′-剪接位点的非标记 RNA 分子则无竞争抑制效应（对比泳道 1 和 5 的结果）。为证明 SC35 是限制因子，Fu 先将标记 RNA 与 SC35 混合预孵育，然后再加入核提取物及竞争性 RNA。比较泳道 4 与 6 发现，在剪接底物与 SC35 混合预孵育条件下，即使存在着竞争者，剪接反应仍能正常发生，因此，SC35 能够保证剪接反应的发生。相似的实验证明，即使以全长 β-球蛋白的 pre-mRNA 作为竞争者，这种定向性依然存在。实验所用 SC35 是昆虫细胞一个克隆基因的产物，因此不可能有其他剪接因子的存在。由此可见，似乎 SC35 单独就足以引发剪接反应的定向。进一步实验研究的结果揭示了定向性存在的必要条件，定向性形成十分快速（1min 内），甚至在冰浴、无 ATP、无 Mg^{2+} 等合理水平上也能形成。

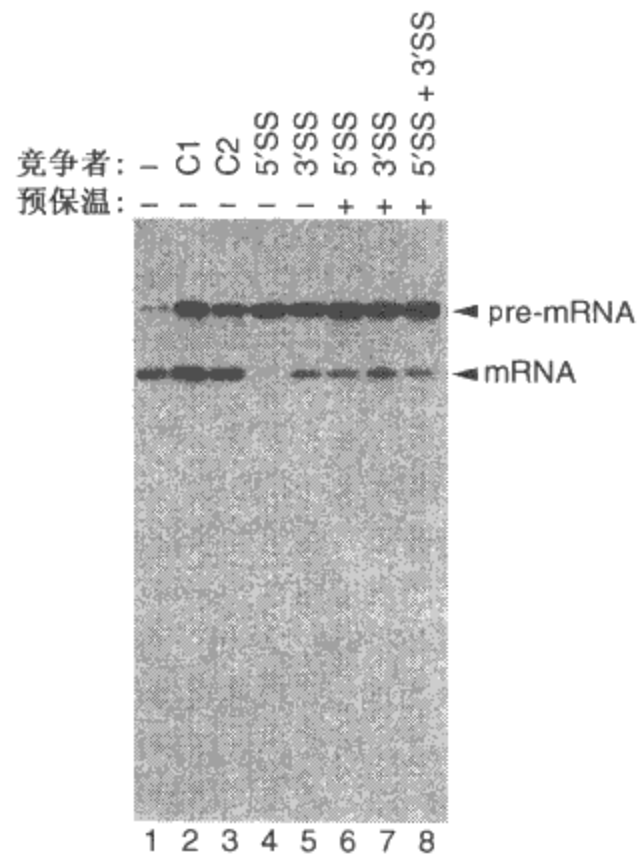


图 14.30 人类 β-球蛋白 pre-mRNA 的定向。 Xiang-Dong Fu 对定向做了如下的竞争实验。将标记的人 β-球蛋白 pre-mRNA 在有 SC35（＋）或无 SC35（－）时孵育，然后加入有或无竞争 RNA 的核提取物，无竞争者的实验组用（－）表示。C1 和 C2 为非特异 RNA，不会干扰剪接。具有 5′-和 3′-剪接位点的 RNA 分别用 5′-SS 和 3′-SS 表示。剪接 2h 后，Fu 电泳标记的 RNA，并放射自显影。pre-mRNA 和成熟 mRNA 的位置在图右侧标注，SC35 能引起定向。（*Source*：Fu，X.-D. Specific commitment of different pre-mRNAs to splicing by single SR proteins. *Nature* 362（2 Sept 1933）f. 1，p. 83. Copyright © Macmillan Magazines Ltd.）

SC35 蛋白是 RNA 结合蛋白即 SR 蛋白家族的一员，具有富含丝氨酸（S）和精氨酸（R）的结构域。因此，Fu 用相同的定向实验测试了其他 SR 蛋白和 RNA 结合蛋白（hnRNP 蛋白）的效应。在所研究的蛋白质中，SC35 的定向性最强，其次分别为

SF2（又称 ASF）和 SRp55；未检测到 SRp20 和 hnRNP A1 的活性；hnRNP C1 和 PTB（又称 hnRNP 1）实际上具有抑制剪接的活性。由此可见，SC35 的定向性是特异的，并不是由一般性 RNA 结合蛋白的功能衍生而来的。

作为定向专一性的进一步证据，Fu 对另一种不同的剪接底物，即人类免疫缺陷性病毒（HIV）的 tat pre-mRNA 进行了研究，已有报道认为 SF2/ASF 能够促进 tat pre-mRNA 的剪接。图 14.31 表明 SF2/ASF 可引起 tat pre-mRNA 的剪接定向。此外，Fu 还比较了 SF2/ASF 和 SC35 定向 tat pre-mRNA 剪接的活性，结果发现，只有 SF2/ASF 能够引起 HIV tat pre-mRNA 的定向剪接，而 SC35 无定向活性，这说明为定向不同 pre-mRNA 的剪接需要不同的剪接定向因子。

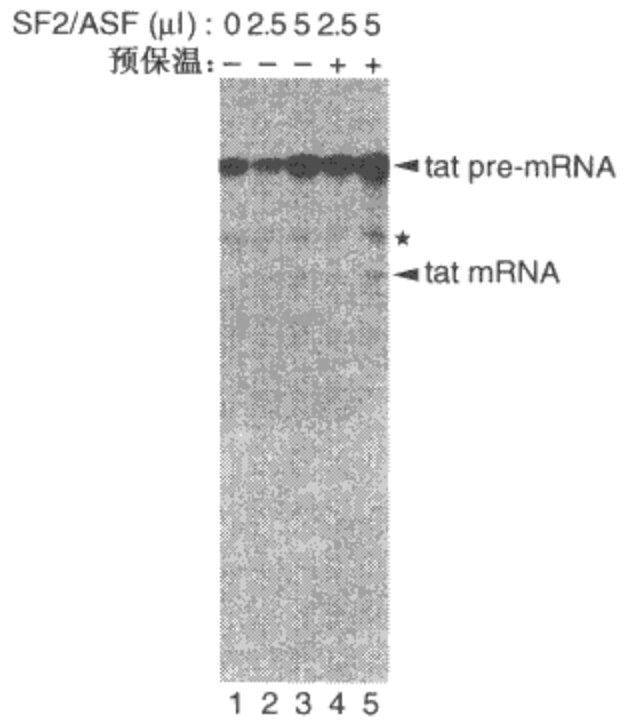


图 14.31　几类 RNA 结合蛋白的定向活性：SF2/ASF 对 tat pre-mRNA 的定向效应。Fu 用不同浓度的 SF2/ASF 进行定向性分析，SF2/ASF 的浓度标注在图上方，pre-mRNA 与剪接因子未预孵育（泳道 1～3）或预孵育（泳道 4 和 5）。对比泳道 5 和 3 可以清楚地看出 SF2/ASF 的定向效应。星号指示的条带是 tat pre-mRNA 的人为降解产物。(*Source*: Fu, X.-D., Specific commitment of different pre-mRNA to splicing by single SR proteins. *Nature* 365(2 Sept 1993) f. 3, p. 84. Copyright © Macmillan Magazines Ltd.)

至今，我们对定向机制还不甚了解，尽管已经知道 U1 能够与定向复合体结合。James Manley 及同事通过凝胶迁移变换实验测定了 U1 snRNP 与标记 pre-RNA 形成稳定复合体的情况。向标记 pre-RNA 中分别加入 U1 或 SF2/ASF 时，并没有复合体的形成；但同时加入这两类蛋白质后则形成了复合体。而且，SF2/ASF 会先行与标记 pre-RNA 结合，因为研究者是按照次序加入蛋白质，其间进行一次洗脱的，所以必须先加入 SF2/ASF 才能使形成复合体。

但是，如果 U1 snRNP 与 pre-RNA 5′-剪接位点的结合依赖于 SF2/ASF，那么，为什么以前的实验结果明显表明 U1 自身单独就可以结合 pre-RNA? 可能的原因就是，先前的实验使用的是细胞核粗提物，自然含有剪接因子。剪接因子与剪接底物之间形成的复合体在先前实验过程中应该是能够检测得到的，但由于实验者主要关注 snRNP 的结合，而没有关注简单蛋白的结合。

小结　RNA 结合蛋白能够确保剪接底物在给定的位点处被剪接。这些 RNA 结合蛋白与剪接底物结合后，可能会招募其他剪接体组分的结合，首先是 U1 的结合。例如，SR 蛋白 SC35 和 SF2/ASF 分别定向人 β-球蛋白 pre-mRNA 和 HIV tat pre-mRNA 的剪接。至少在某些情况下，这种定向效应需要 U1 的参与。

桥接蛋白与定向　有关剪接定向事件的一个额外插曲就是酵母细胞没有 SR 蛋白，而酵母细胞是最初开展剪接体循环实验的研究系统。这一发现暗示，酵母细胞的定向机制可能不同于哺乳动物。但随后的研究表明，酵母和哺乳动物的定向复合体具有许多共同的特征。我们先研究酵母定向复合体中参与将内含子的 5′端和 3′端桥接在一起的蛋白质，然后将其与哺乳动物的相应蛋白质做一对比。

1993 年，Michael Rosbash 及同事为发现编码参与构成酵母定向复合体的蛋白质基因而开展了相关研究工作。由于 U1 snRNA 是主要的也是最先参与定向复合体形成的组分，因此研究者决定寻找编码与 U1 snRNA 相互作用的蛋白质的基因。为发现这些基因，他们采用了**合成致死筛选法**（synthetic lethal screen）：首先在编码 U1 snRNA 的基因内引入温度敏感型突变，突变体 U1 snRNA 在较低温度（30℃）下有功能活性，而在较高的温度（37℃）下则丧失功能活性。研究者推论携带这种突变的菌株对与 snRNA 相互作用蛋白质的突变会十分敏感，第二次突变会产生致死效应，即使在较低温度条件下，双突变酵母菌株也很难存活。所以将这种突变体称为“突变致死”（mutant-u-die），简写为 Mud。在野生型酵母细胞内单独发生第二种突变不会产生致死效应，但如果酵母细胞已发生了第一种突变，那么再发生第二类突变就会产生致死效应。从这个意义上讲，这两类突变的致死效应是相辅相成的，即一种突变的致死效应是以另一种突变的发生为前提条件的。利用这种突变策略，发现了编码 Mud2p 蛋白

的基因。

随后的研究显示，Mud2p 的功能发挥有赖于邻近内含子 3′端、位于套索分支点处的一个自然序列。这暗示 Mud2p 不仅与结合在内含子 5′端的 U1 snRNP 相互作用，还与邻近内含子 3′端处其他物质相互作用。但一个主要的问题是：与内含子 5′端和 3′端发生的相互作用仅仅依靠 Mud2p 自身就能够完成，还是需要其他因子的参与？1997 年，Nadja Abovich 和 Rosbash 通过合成致死筛选实验回答了这个问题。研究者在 *MUD2* 基因内引入突变，然后筛选能使 *MUD2* 突变体产生致死表型的第二个突变，但该突变不会使野生型产生致死表型。通过综合致死筛选策略鉴定出了 *MSL-5*（Mud synthetic lethal -5）基因，该基因编码的蛋白质产物最初被命名为 Msl5p，但在其结合特性确定后，被重新命名为**分支点桥接蛋白**（branchpoint bridging protein，BBP）。

Abovich 和 Rosbash 推测，BBP 既可与结合在 5′端的 U1 snRNP 相互作用，也可与结合 3′端的 Mud2p 蛋白相互作用，从而将内含子的 5′端与 3′端桥接在一起。为验证这一假设，研究者利用包括**酵母双杂交实验**（yeast two-hybrid assay）在内的综合研究策略阐释了桥接蛋白间的相互作用关系。图 14.32 描述的酵母双杂交是一项十分灵敏的遗传学实验技术，可检测两个蛋白质间的结合关系，即使是瞬间的结合。酵母双杂交实验的设计原理基于以下两个事实（已在第 12 章讨论过）：①典型的转录激活子具有 DNA 结合域和转录激活域；②这两个功能域具有独立的活性（self-contained activity）。为分析 X、Y 这两个蛋白质之间的结合关系，可以使这两个蛋白质在酵母细胞内以融合蛋白的形式表达，如图 14.32b 所述，将蛋白质 X 与 DNA 结合域融合、蛋白质 Y 与转录激活域融合。如果蛋白质 X 和 Y 能相互作用，那么，就可将 DNA 结合域和转录激活域结合在一起，从而激活报告基因的转录（*lacZ*）。

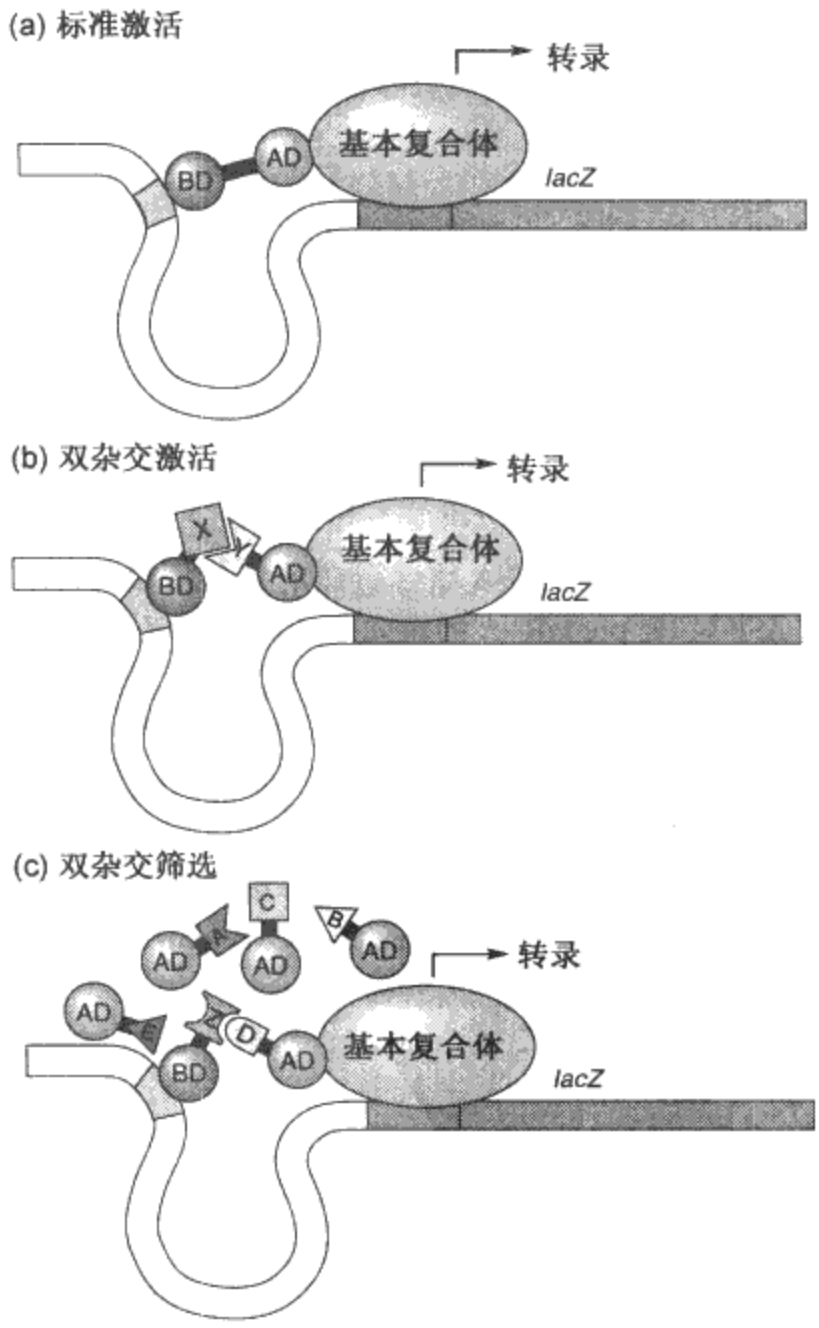


图 14.32　酵母双杂交原理。（a）转录激活的标准模型。激活子的 DNA 结合域（BD，红色）结合增强子（紫色），激活域（AD，绿色）与基本复合体（橙色）相互作用，招募到启动子（红褐色）上从而激活转录。（b）蛋白质-蛋白质相互作用的酵母双杂交实验。将 X 蛋白（蓝绿色）连接到 DNA 结合域上形成杂合蛋白；将另一个蛋白质（Y，黄色）连接到激活域上形成第二个杂合蛋白；将分别编码这两个杂合蛋白的质粒同时转化同一酵母细胞，此酵母细胞具有合适的启动子、增强子及报告基因（如 *lacZ*，紫色）。这两个杂合蛋白可以结合在一起而发挥激活子的作用，激活转录后，产生大量报告基因编码的蛋白质，可利用显色法检测，如 X-gal 。一种杂合蛋白提供 DNA 结合域，另一种蛋白质提供转录激活域，通过蛋白 X 和 Y 之间的相互作用而将激活子的这两部分结合在一起，如果蛋白 X 和 Y 之间的无相互作用关系，则无激活子的形成，进而也就不能激活报告基因的表达。GAL4 蛋白的 DNA 结合域和转录激活域常用于酵母双杂交试验。但也有其他的，如 Abovich 和 Rosbash 使用的结合域和激活域。（c）利用**酵母双杂交筛选**（Yeast-two-hybrid screen）与 Z 蛋白相互作用的蛋白质。用两种质粒转化酵母细胞，一种质粒编码结合域（红色）及与之偶联的诱饵蛋白 Z（蓝绿色），另一个质粒为一套质粒中的一个，这套质粒库中的每一个质粒整合有一 cDNA 片段及激活域的编码区，能够编码由激活域（绿色）和未知的 cDNA 表达产物（靶蛋白）的融合蛋白。为便于说明，图中显示不同质粒表达的融合蛋白产物。靶蛋白 D（黄色）能够与诱饵蛋白 Z 相互作用，从而将 DNA 结合域和转录激活域结合在一起，激活报告基因的表达。从阳性克隆菌株中提取质粒，通过相应的分析即可获知有关靶蛋白的相关信息。

利用酵母双杂交系统可以找到与已知蛋白质(Z)相互作用的未知蛋白质。按照图 14.32c 所示的筛选程序，首先将 cDNA 文库中的 cDNA 片段与转录激活域的编码区融合在一起，然后使这些杂合基因以及编码 DNA 结合域-Z 融合蛋白的杂合基因在酵母细胞内共表达。在实际操作中，每个酵母细胞只能表达一个不同的融合蛋白（如 AD-A、AD-B、AD-C 等），同时还表达 BD-Z 融合蛋白，但为简便起见，这里将所有的融合蛋白置于一起而阐述。从图示中可以看到，AD-D 结合了 BD-Z，激活报告基因的转录，但其他融合蛋白由于不能与 BD-Z 相互作用，因而也就不能激活转录基因。一旦筛选到激活转录的克隆，就可以分离携带 AD-D 杂合基因的质粒，通过测序获知 D 序列的编码信息。由于酵母双杂交分析是一种间接检测蛋白质间相互作用的方法，可能会有假阳性出现。因此，依此鉴定出的蛋白质-蛋白质相互作用还必须用直接的分析方法加以验证。

由于 Abovich 和 Rosbash 已经事先了解了蛋白质间可能的相互作用。因此，他们构建了含有目的蛋白基因以及 DNA 结合域或转录激活域基因的质粒，然后将这些重组质粒两两组对，转化酵母细胞。例如，其中的一个实验，用于转化酵母细胞的重组质粒为一个重组质粒能够表达 LexA BD-BBP 杂合蛋白；另一个重组质粒能够表达 B42 AD-Mud2p 杂合蛋白。图 14.33a（第 1 列、第 1 行）的结果表明，转化有这两个重组质粒的酵母细胞能够激活 *lacZ* 基因的表达，在 X-gal 指示剂的平皿上产生的印迹可以证实 *lacZ* 基因的表达。这个分析实验结果表明 BBP 可与 Mud2p 结合。图 14.33a（第 1 列、第 2 行）显示，BBP 也可与 U1 snRNP 的多肽组分 Prp40p 结合。但 Mud2p 不能与 Prp40p 结合。所以，BBP 可在 U1 snRNP 与 Mud2p 之间发挥桥接分子的作用，其中 U1 snRNP 结合在内含子 5′端，而 Mud2p 与邻近内含子 3′端的分支点结合。BBP 以这种方式将剪接底物中内含子界定出来，并将内含子的两个末端桥接在一起便于剪接。由于已经预知 Prp8p 可与 Prp40p 结合，所以，Abovich 和 Rosbash 将 Prp8p 蛋白作为实验的阳性对照。图 14.33b 将通过将通过酵母双杂交实验检测到的蛋白质-蛋白质间的相互作用进行了归纳总结。研究者发现，束缚在琼脂糖珠上的 BBP 能够与 Prp40p 和 Mud2p 发生共沉淀，从而证实了酵母双杂交的实验结果。

Abovich 和 Rosbash 注意到，酵母的 Mud2p 和 BBP 蛋白分别与哺乳动物的 U2AF65 和 SF1 蛋白相似。如果这两个哺乳动物蛋白质与它们的酵母对应

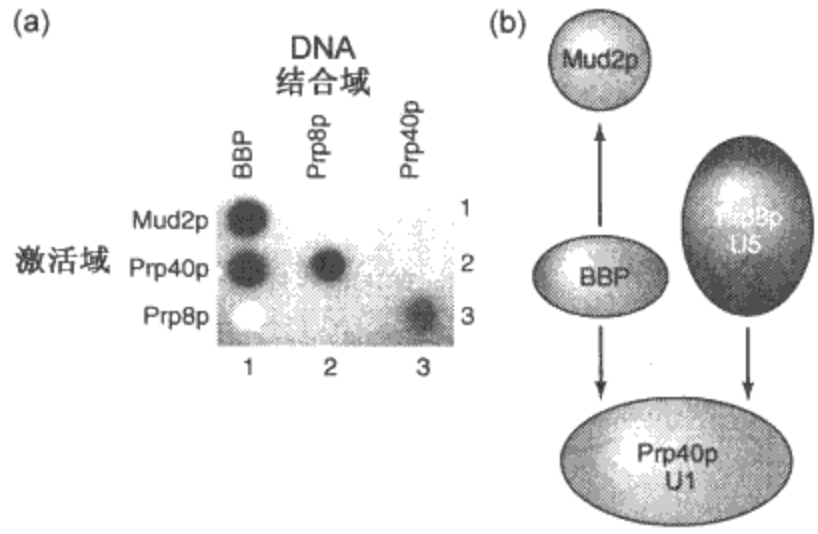


图 14.33　BBP 与其他蛋白质相互作用的酵母双杂交分析。（a）实验结果。与 DNA 结合域连接的蛋白质在图上方标出；与转录激活域连接的蛋白质在图左侧标出。Abovich 和 Rosbash 将转化有双质粒的酵母细胞培养在涂布有 X-gal 的培养皿上，通过菌斑的显色来衡量报告基因 *lacZ* 的激活情况，呈暗色的菌株表明报告基因被激活。例如，第 1 列第 1 行和第 2 行的两个暗色菌斑，表明 BBP 可分别与 Mud2p 和 Prp40p（U1 snRNP 的组分）相互作用。另一个阳性反应表明 Prp40p 与 Prp8p（U5 snRNP 的组分）有相互作用。（b）实验结果总结 。示意了图 a 酵母双杂交实验所揭示的蛋白质-蛋白质间相互作用关系。(*Source*: Abovich N. and M. Rosbash, Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals *Cell* 89 (2 May 1997) f. 5, p. 406. Reprinted by permission of Elsevier Science.)

蛋白具有相似的功能特征，那么 U2AF65 与 SF1 就应彼此结合。为验证这个假设，研究者同样利用酵母双杂交和共沉淀方法进行研究。结果发现 U2AF65 和 SF1 确实存在相互作用关系。图 14.34 总结了 BBP 与其哺乳动物同源蛋白 SF1（又称为 mBBP）的桥接作用。在讨论选择 3′-剪接位点的内容时，已经了解了 U2AF 的功能特征，U2AF65 结合在 3′-剪接位点上游的多聚嘧啶区，而 U2AF35 与 3′-剪接位点的 AG 结合。在图 14.34 中，U2AF65 和 U2AF35 以 U2AF 的形式共同结合在 3′-剪接位点上。

Rosbash 研究小组通过进一步的深入研究证明，BBP 也能识别分支点序列 UACUACC，在定向复合体中能与该序列结合（或十分靠近该序列）。所以，BBP 也是一类 RNA 结合蛋白，而且 BBP 现在的含义是“分支点结合蛋白”。

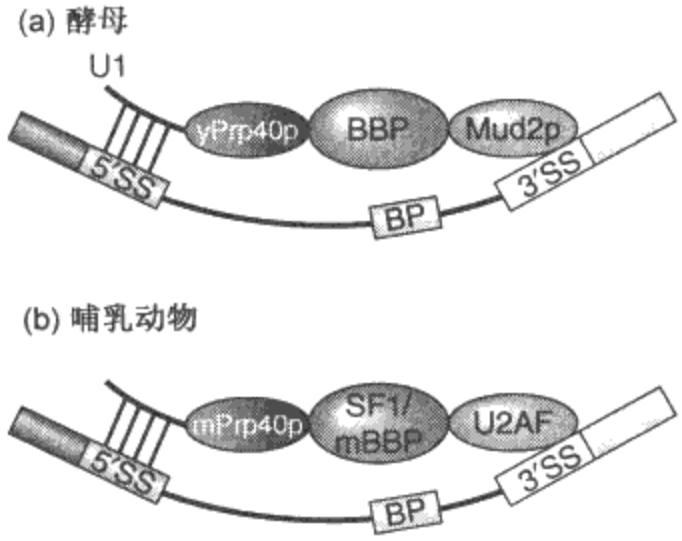


图 14.34　在酵母（a）和哺乳动物（b）细胞内，内含子-桥接蛋白-蛋白质相互作用概括。5′-SS 是 5′-剪接信号；BP 为分支点；3′-SS 是 3′-剪接信号。除 Prp8p 与 Prp40p 的相互作用外，其他蛋白质间的相互作用均由 Abovich 和 Rosbash 证实。（*Source*：Adapted from Abovich, N. and M. Rosbash, Cross-intron bridging interactions in yeast commitment complex are conserved in mammals. *Cell* 89：409，1997.）

小结　在酵母定向复合体中，分支点桥接蛋白（BBP）能够与内含子 5′端的 U1 snRNP 和邻近内含子 3′端的 Mud2p 发生结合。此外，BBP 还能与内含子 3′端附近的 RNA 序列结合。因此，BBP 能够将内含子的两个末端桥接在一起，在剪接之前将内含子界定出来。哺乳动物的 BBP 同源蛋白 SF1（mBBP）在哺乳动物定向复合体中具有相同的桥接作用。

RNA 聚合酶 II CTD 的功能　如本章开始所提及的，同 5′端加帽、3′端多聚腺苷酸化一样，pre-mRNA 的剪接同样受 RNA 聚合酶 II 的最大亚基 Rpbl 的 C 端功能域（CTD）的调控。我们如何了解 CTD 在剪接过程中所发挥的功能？2000 年，Changqing Zeng 和 Susan Berget 利用标记的剪接底物进行体外剪接反应，该剪接底物含有两个完整的外显子和一个位于两个外显子之间的内含子，其结构如图 14.35b 上方的简图所示。为确保反应的发生，研究者同时向反应体系中分别添加了 CTD-GST（谷胱甘肽-S-转移酶）或 GST。

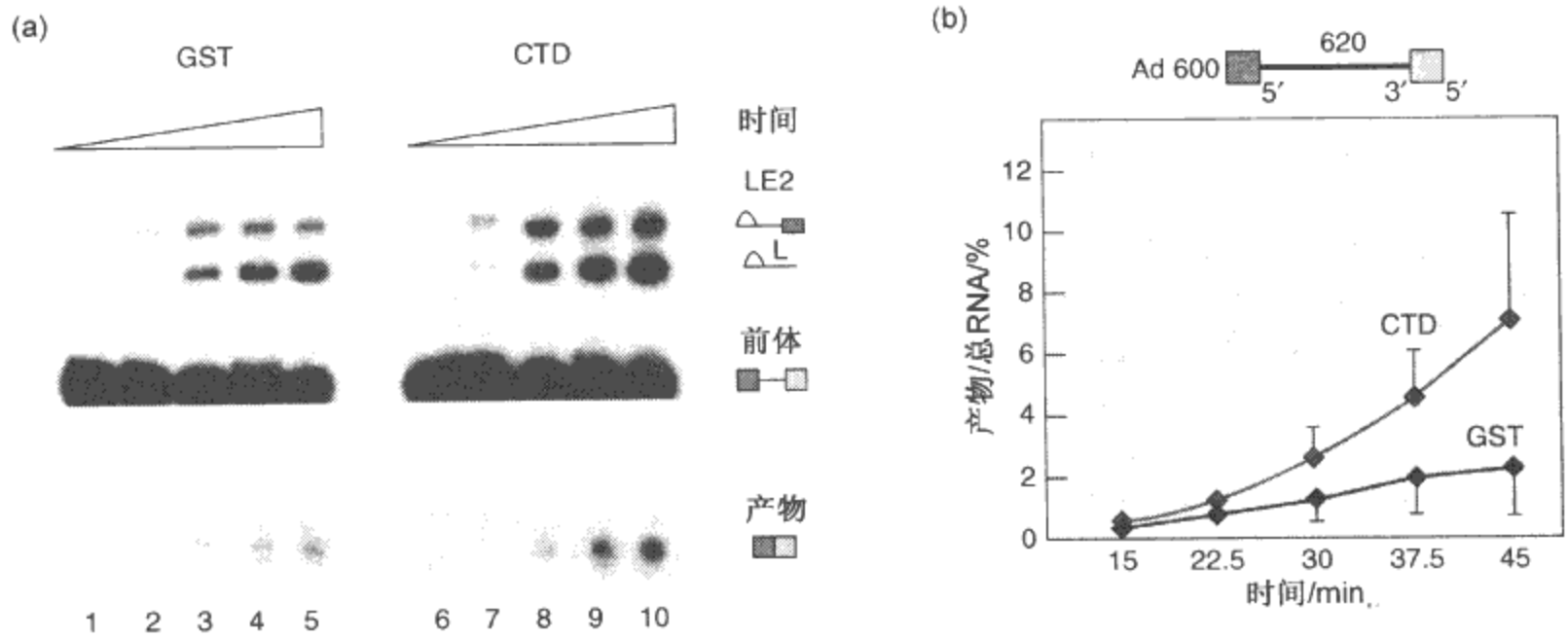


图 14.35　CTD-GST 能够促进体外剪接反应。（a）剪接反应。Zeng 和 Berget 将 ^{32}P-标记的剪接底物（Ad600）与添加有 GST（左）或 CTD-GST（右）的剪接提取物共同孵育，剪接底物的结构见图（b）的上方，上方的楔形符号表示孵育时间的延长。然后电泳提取物，分离前体、中间体和最终产物。图左侧标出了 RNA 的位置，并绘出简图以助识别。CTD 使反应效率提高 3～5 倍。（b）实验结果的曲线图。根据 RNA 产物的百分率与时间（min）关系作图，蓝色曲线为添加 GST 的剪接反应结果；红色曲线为添加 CTD-GST 的剪接反应结果。（*Source*：Copyright © American Society for Microbiology，*Molecular and Cellular Biology* vol. 20，No. 21，p. 8294. fig. 1，2000.）

图 14.35a 显示，通过测量套索状外显子中间体、套索状内含子及外显子产物的数量可以看出，融合蛋白 CTD-GST 能促进剪接。与 GST 相比，CTD-GST 促进剪接的程度是 GST 的 3～5 倍，当然 GST 不具有促进剪接的效应。值得注意的是，两组实验的剪接中间体和最终产物出现的时间没有差异，但添加 CTD-GST 融合蛋白的实验组在相同时间内能够积累更多得剪接中间体分子和最终产物，这说明 CTD 能够招募剪接底物结合到有活性的剪接复合体上。

令人感兴趣的是，CTD-GST 不能促进含有不完整外显子的底物的剪接，图 14.36 显示了这一现象。剪接底物 Ad 100 和 MT16-L 只含有 1 个完整的外显子，但 CTD-GST 能促进这些底物的剪接。当底物 MT16-S 含有 2 个完整外显子和 1 个不完整外显子时，CTD-GST 不会对这个底物的剪接产生任何影响。在相似的实验中，含有 1 个完整外显子和 1 个不完整外显子的底物其剪接反应也不受 CTD 促进。

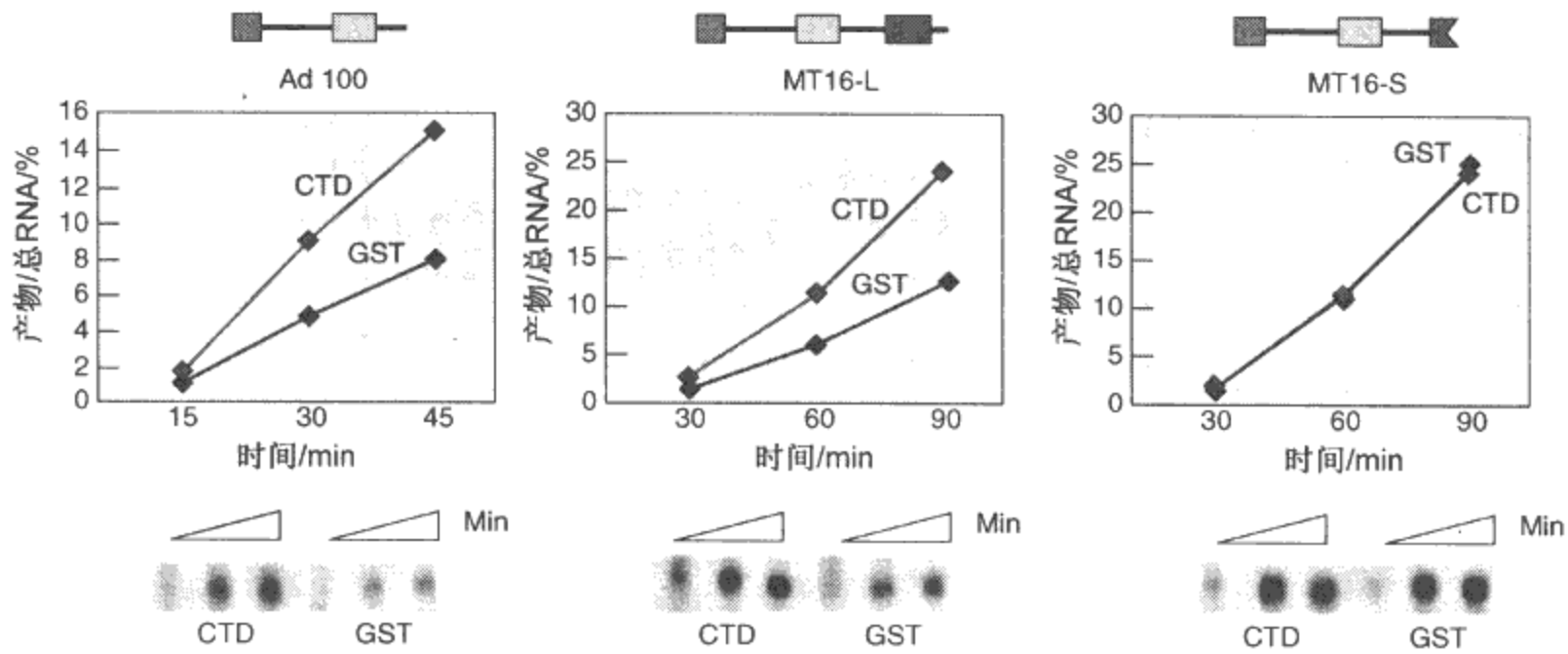


图 14.36　在外显子界定或内含子界定条件下，CTD-GST 对剪接反应的影响。Zeng 和 Berget 利用图 14.35 的方法，对图上方所示的 3 种标记剪接底物进行体外剪接。前两个底物有完整的外显子，按外显子界定模型进行剪接；第三个底物 MT16-S 有一个不完整的外显子，按内含子界定模型进行剪接；凝胶电泳结果见图底部，实验结果如上方曲线图所示。蓝色曲线为添加 GST 的剪接反应结果；红色曲线为添加 CTD-GST 的剪接反应结果。(*Source*: Copyright © American Society for Microbiology, *Molecular and Cellular Biology* vol. 20, No. 21, p. 8294, fig. 4, 2000.)

先前的实验已表明，CTD 能够与 snRNP 和 SR 蛋白结合。据此，Zeng 和 Berget 提出，在 RNA 聚合酶合成外显子后，CTD 通过装配剪接因子到外显子上而促进剪接反应（图 14.37）。但为什么只有在剪接底物具有完整外显子序列时才能表现出这一功能？Changqing Zeng 和 Susan Berget 从**外显子界定**（exon definition）方面解释了这些结果。基于其他一些实验，Berget 及同事提出了给定 pre-mRNA 的两种剪接方式，即**外显子界定**和**内含子界定**（intron definition）。外显子界定是指剪接因子通过识别外显子的末端而切除位于外显子之间的内含子；内含子界定是指剪接因子通过识别内含子末端而不是外显子末端而切除内含子。对外显子界定而言，要求所有的外显子序列必须完整，这样就能够准确判定某个序列是否为真正的外显子，否则就会因外显子的不确定而不能发生外显子界定的剪接反应。但内含子界定的剪接反应仍能正常进行。如果这一假说是正确的话，则很明显，CTD 是不会促进内含子界定的剪接反应。

免疫剥离实验的结果进一步证实了 CTD 在外显子界定模型中发挥作用的假说。Zeng 和 Berget 通过免疫剥离从提取物中去除 RNA 聚合酶 II，结果发现，由于 RNA 聚合酶 II 的部分缺失，依赖于外显子界定的底物剪接受到了抑制，但几乎不影响依赖于内含子界定的底物剪接。向经免疫剥离处理的提取物中重新添加 CTD 可恢复依赖于外显子界定的底物剪接。

小结　RNA 聚合酶 II Rpbl 亚基的 C 端功能域（CTD）能够促进外显子界定的底物剪接，但不会促进内含子界定的底物剪接。CTD 与剪接因子结合，从而将剪接因子装配到外显子的两个末端，引发剪接反应。

选择性剪接　通过前面对剪接定向性的讨论，自然而然地引发出另一个重要的话题：**选择性剪接**（alternative splicing）。真核生物大约有 5% 的 pre-mRNA 可按一种以上的方式进行剪接，产生两条或多条编码不同蛋白质的 mRNA 分子。人类 60% 的转录物可发生选择性剪接。从一种剪接方式向另一种剪接方式的转换无疑会涉及剪接的定向性，在本

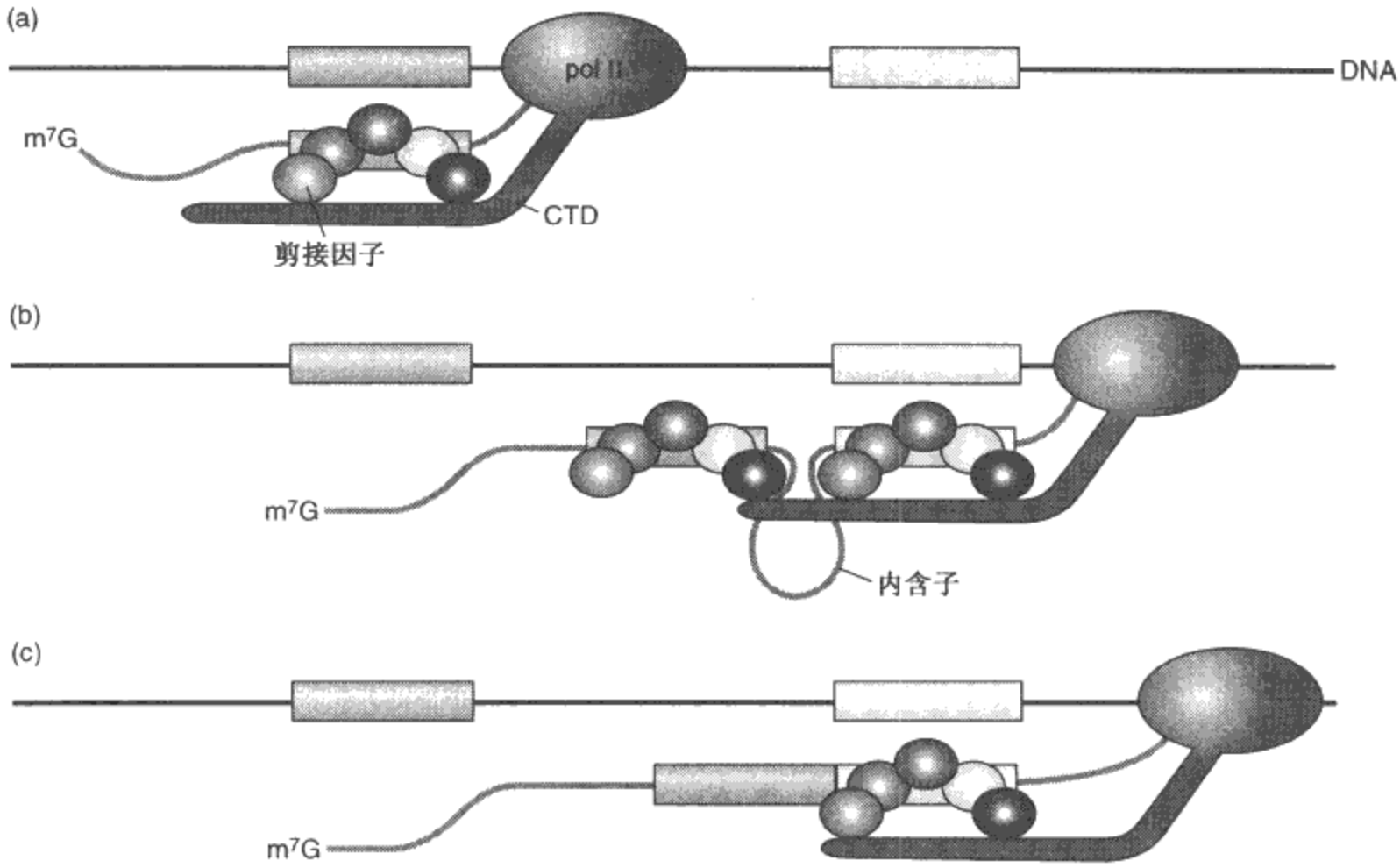


图 14.37　CTD 参与外显子界定的模型。（a）RNA 聚合酶转录出第一个外显子后，CTD 介导剪接因子装配到 pre-mRNA 外显子的两个末端，从而界定了这个外显子。（b）RNA 聚合酶转录出第二个外显子，CTD 以同样方式介导该外显子的界定。CTD 也能将两个外显子彼此拉近，为剪接做好准备。（c）当 RNA 聚合酶还在继续转录时，两个外显子已被拼接。(*Source*: Adapted from Zeng, C. and S. Berget, Participation of the C-terminal domain of RNA polymerase II in exon definition during per-mRNA splicing. *Molecular and Cellular Biology* 20 (2000) p. 8299. F. 9.)

节最后我们再讨论这个话题。

1980 年，David Baltimore 及同事首次在小鼠免疫球蛋白 μ 重链基因中发现了选择性剪接的事例。小鼠球蛋白 μ 重链有分泌型（μ_s）和膜结合型（μ_m）两种形式，其区别在于，膜结合型的 C 端具有能使之锚定在膜上的一个疏水区，而分泌型的 C 端无此疏水功能区。通过核酸杂交分析，Baltimore 及同事发现，分别编码这两种蛋白质的 mRNA 分子的 5′端序列相同，但 3′端序列完全不同。研究者克隆了编码免疫球蛋白 μ 重链固定区的胚系基因（C_μ 基因），发现该基因既能编码分泌型重链（μ_s）的 3′区，也能编码膜结合型重链（μ_m）的 3′区，只是这两个序列分属两个独立的外显子。由此可见，同一条 pre-mRNA 分子通过两种不同的剪接方式，可形成了分别编码 μ_s 和 μ_m 的两条成熟的 mRNA 分子，如图 14.38 所示。因此，选择性剪接能够决定一个基因蛋白质产物的性质，因而可调控基因的表达。

选择性剪接具有重要的生物学功能。果蝇的性别决定系统是选择性剪接调控基因表达的最好例证，果蝇的性别决定涉及 *sxl*（性别致死，*sex lethal*）、*tra*（转换，*transformer*）、*dsx*（双性别，*doublesex*）3 个基因 pre-mRNA 的选择性剪接。图 14.39 总结了果蝇基因选择性剪接的模式。这些基因的转录物若进行雄性特异性剪接，则导致雄性个体的发育；这些基因的转录物若进行雌性特异性剪接，则导致雌性个体的发育。

而且，这些基因在级联剪接反应中有如下功能。*sxl* 基因转录物的雌性特异性剪接，导致活性蛋白质产物的形成，此产物能进一步增强 *sxl* 基因转录物的雌性特异性剪接反应，同时也引发 *tra* 基因转录物的雌性特异性剪接反应，产生有活性的 Tra 蛋白（实际上，即使在雌性个体内，大约也有一半量的 *tra* 基因转录物进行雄性特异性剪接反应，但只能产生无活性的蛋白质产物，因此，雌性特异性剪接占主导地位）。活性 tra 蛋白与 tra-2 蛋白协同作用，引发 *dsx* 基因转录物的雌性特异性剪接反应，其蛋白质产物能抑制雄性特异基因的表达，从而导致雌性个体的发育。

相反，*sxl* 转录物的雄性特异性剪接产生无活性的蛋白质产物，因为剪接形成的 mRNA 分子内有一个外显子含有终止密码子。所以 *tra* 转录物可进行

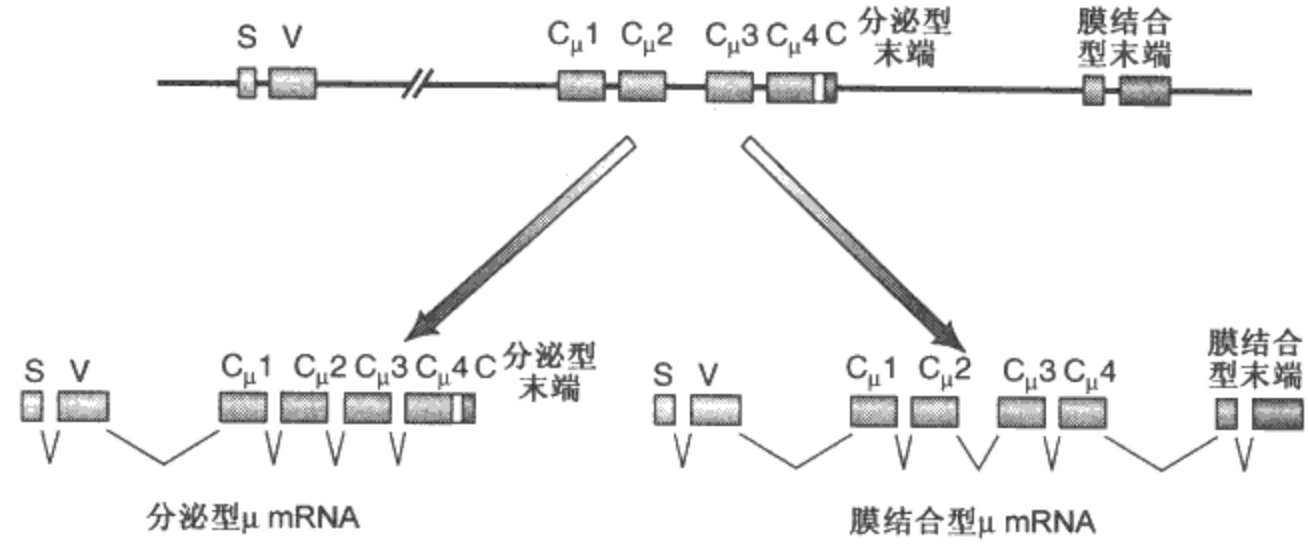


图 14.38　小鼠免疫球蛋白 μ 重链基因的选择性剪接模式。图上方为基因结构图示。条框代表外显子；S 外显子（紫色）编码将蛋白质产物分泌到质膜或胞外的信号肽；V 外显子（橙色）编码蛋白质的可变区；C 外显子（蓝色）编码蛋白质的恒定区。μ_s 蛋白分泌末端的编码区（黄色）邻近第 4 个恒定外显子（Cμ4）的末端，其后是一小段非翻译区（红色）及一较长的内含子。最后是两个外显子，其中第一个外显子（绿色）编码 μ_m mRNA 的膜锚定区，第二个外显子（红色）是 μ_m mRNA 末端的非翻译区。指向左、右两侧的箭头指示剪接的两种模式，产生免疫球蛋白 μ 重链的两种类型：分泌型 μ_s 和膜结合型 μ_m。(*Source*: Adapted from Early P., J. Rogers, M. Davis, K. Calame, M. Bond, R. Wall, and L. Hood, Two mRNAs can be produced from a single immunoglobulin γ gene by alternative RNA processing pathways. *Cell* 20: 318, 1980.)

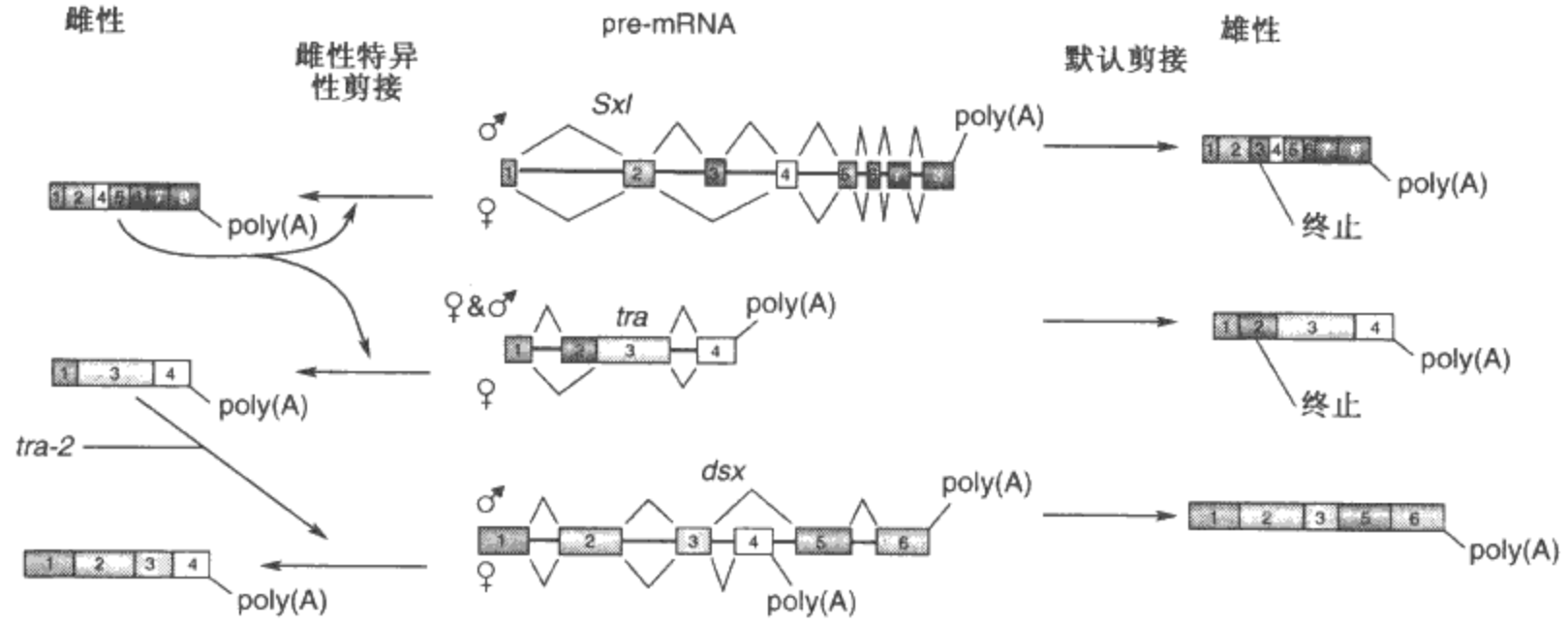


图 14.39　果蝇性别决定中选择性剪接的级联反应。*sxl*、*tra* 和 *dsx* 基因 pre-mRNA 的结构在雌雄果蝇个体中是一致的，见图中部所示。雌性特异剪接模式和雄性特异剪接模式分别在结构图的上方和下方标出。*sxl* pre-mRNA 的雌性特异性剪接涉及 1、2、4～8 等 7 个外显子的拼接，而其雄性特异性剪接涉及全部 8 个外显子的拼接，其中外显子 3 含有终止密码子。这意味着该转录物的雄性特异性剪接产生短的无活性蛋白质。同样，*tra* pre-mRNA 的雌性特异性剪接涉及 1、3、4，导致功能活性蛋白质的产生，而其雄性特异性剪接涉及全部 4 个外显子的拼接，其中外显子 2 含有终止密码子的，产生无活性蛋白质产物。最左侧的长箭头指示基因蛋白产物对剪接有正控效应，即 Sxl 蛋白会引发 *sxl* pre-mRNA 和 *tra* pre-mRNA 的雄性特异性剪接反应。雌性 *tra* 基因产物与 *tra*-2 产物协同作用，引发 *dsx* 基因转录物的雄性特异性剪接。(*Source*: Adapted from Baker, B. S. Sex in flies: The spice of life. *Nature* 340: 523, 1989.)

默认（雄性特异性剪接）剪接，剪接产物因包含终止密码子而不能翻译出功能 Tra 蛋白。由于没有 *tra* 基因的蛋白质产物，那么，处于发育期的细胞将按默认的雄性特异性剪接模式对 *dsx* 转录物进行剪接，

其产物将使雌性特异性基因失活，从而导致雄性个体的发育。

如何控制选择性剪接？在对剪接定向性机制了解的基础上，我们猜测，RNA 结合剪接因子可能参与对选择性剪接的调控。确实，由于 *sxl* 和 *tra* 产物能够分别决定 *tra* 基因转录物和 *dsx* 基因转录物的剪接位点，所以我们推测这些蛋白质应该是剪接因子，能够确保雌性特异性剪接模式的发生，根据这一假设，Sxl 蛋白和 Tra 蛋白应该均为 SR 蛋白。

为进一步阐明剪接位点选择的机制，Tom Maniatis 及同事着重研究了 Tra 蛋白和 Tra-2 蛋白（分别为 *tra* 和 *tra-2* 基因的蛋白质产物）对 *dsx* pre-RNA 的雌性特异性剪接。他们发现，Tra 和 Tra-2 这两个蛋白质通过结合在 *dxs* pre-mRNA 的雌性特异性 3′-剪接位点下游约 300nt 处的调控区而发挥作用，该调控区含 6 个 13nt 序列的重复元件（repeat element）。

Tra 和 Tra-2 是 *dxs* per-mRNA 进行雌性特异性剪接所必需的，但是否为充分条件？为阐明事实，Ming Tian 和 Maniatis 针对剪接的定向性，开展了相关的研究工作。研究者以只含外显子 3 和 4 及两者之间内含子的标记截短 *dsx* pre-mRNA 为体外剪接底物，该底物在体外可被剪接。然后加入 Tra、Tra-2 和**经微球菌核酸酶**（micrococcal nuclease，**MNase**）处理后的细胞核提取物，提供除 Tra、Tra-2以外的其他蛋白质，以用于定向复合体的形成。微球菌核酸酶能够降解 snRNA 而保留了 snRNP 中的蛋白质组分。然后在加入未经核酸酶处理的细胞核提取物及过量的竞争 RNA。如果定向复合体在预孵育阶段就已形成，则标记的 pre-mRNA 将会被剪接；如果在预孵育阶段定向复合体没有形成，那么竞争 RNA 就会阻断标记 pre-mRNA 的剪接。为分析标记底物的剪接情况，Tian 和 Maniatis 对反应产物进行凝胶电泳，通过放射自显影确定 RNA 的种类。结果发现，在无 MNase 处理的核提取物存在时，Tra 和 Tra-2 单独存在不足以促使定向复合体的形成。因此，核提取物中存在着某些能补偿 Tra 和 Tra-2 功能的蛋白质，可引发定向复合体的形成。

为确定其他必需的剪接因子，Tian 和 Maniatis 首先纯化了所有的 SR 蛋白，发现 SR 蛋白混合物能够补偿 Tra 和 Tra-2 的功能。接下来，研究者获得 4 个纯化重组的 SR 蛋白和 2 个高度纯化的非重组 SR 蛋白，然后通过剪接定向性实验分析这些 SR 蛋白对 Tra 和 Tra-2 的补偿功能。在该实验中，用纯化的蛋白质取代了前面实验中的 MNase 处理的细胞核提取物。图 14.40 泳道 1 的结果显示，在没有其他任何 SR 蛋白时，Tra 和 Tra-2 不能单独引发剪接反应；泳道 2 显示，经硫酸铵（ammonium sulfate，AS）沉淀获得 SR 蛋白混合物能够补偿 Tra 和 Tra-2 的功能；其他泳道显示重组 SR 蛋白和高度纯化的非重组 SR 蛋白对 Tra 和 Tra-2 的补偿效应，其中有些 SR 蛋白具有功能补偿效应，而有些 SR 蛋白则无功能补偿效应。特别是，SC35、SRp40、SRp55 和 SRp75 能补偿 Tra 和 Tra-2 的功能，但 SRp20 和 SF2/ASF 无功能补偿效应。由此可见，Tra 和 Tra-2 以及外加任何一种有活性的 SR 蛋白就足以确保 *dsx* pre-mRNA 发生雌性特异性剪接。

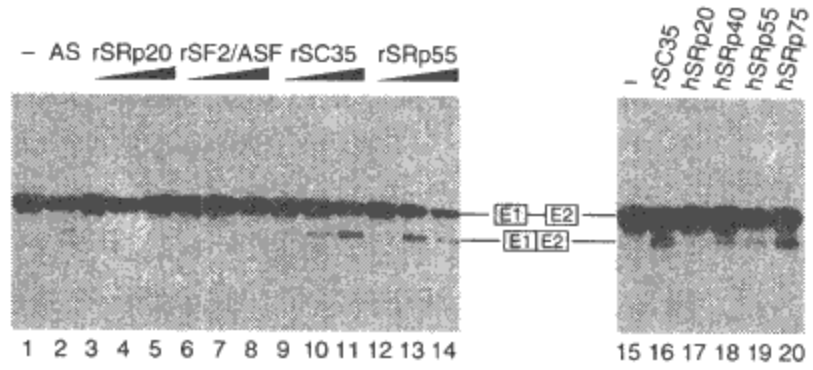


图 14.40 *dsx* pre-mRNA 雌性特异性剪接的定向性分析。Tian 和 Maniatis 通过 *dsx* pre-mRNA 的体外剪接试验，分析了各种 SR 蛋白与 Tra 、Tra-2 的互补能力。泳道 1：无补偿蛋白；泳道 2：含有经硫酸铵（AS）沉淀的 SR 蛋白混合物；泳道 3 ～14：含有不同的 SR 蛋白，SR 蛋白种类在泳道的上方标出；泳道 15：与泳道 1 相同的另一个负对照；泳道 16：含有最大量的重组蛋白 SC35，同泳道 11；泳道 17～20：含有纯化的非重组 SR 蛋白，SR 蛋白种类在泳道的上方标出。在两张放射自显影图像的中间标注了剪接底物（上方条带）及剪接产物（下方条带）泳动的位置。（*Source*：Tian，and M. Maniatis，A splicing enhancer complex controls alternative splicing of *doublesex* pre-mRNA. *Cell* 74（16 July 1993）f. 5. p. 108. Reprinted by permission of Elsevier Science.）

特定剪接模式的发生涉及 SR 蛋白与 pre-mRNA 的结合，已知 Tra 和 Tra-2 结合在重复元件上，其他 SR 蛋白是否也与此元件结合？为探明事实，Tian 和 Maniatis 将含有重复元件的 RNA 连接到树脂上，利用亲和层析分离法，从含有重复元件的 RNA 分子上分离获得了与之结合的蛋白质，电泳后进行免疫印迹（Western 斑点杂交）。最后，分别利用抗 Tra、抗 Tra-2 和抗 SR 的抗体在 3 组独立实验中分析免疫印迹的蛋白质组分。他们检测到了 Tra 蛋白、Tra-2 蛋白以及大量的 SRp40，其中在一个条带中也可能含有 SF2/ASF 或 SC35。由于在剪接定向性分析实验中是 SC35 而不是 SF2/ASF 能补偿 Tra 和

Tra-2 的功能，因此我们推测该条带中的蛋白质组分应该是 SC35。在缺少 Tra 和 Tra-2 蛋白时，SR 蛋白与 RNA 都没有显著的结合。该实验证明，在 Tra 和 Tra-2 蛋白存在时，只有两个 SR 蛋白能够很好地与含有重复元件的 RNA 结合。虽然这并不表明结合与剪接定向之间存在必然联系，但所结合的 2 个 SR 蛋白在定向过程中可补偿 Tra 和 Tra-2 的事实给了我们一定的暗示。

小结 许多真核生物基因的转录物可进行选择性剪接，这对基因的蛋白质产物有着重要的影响。例如，通过选择性剪接可产生分泌型和膜结合型两种不同的蛋白质，甚至能够产生具有活性和无活性之分的蛋白质。在果蝇的性别决定途径中，3 个基因的转录产物就经历了选择性剪接。*tra* 基因转录物经雌性特异性剪接后产生功能活性 Tra 蛋白，引发 *dsx* pre-mRNA 的雌性特异性剪接，最终导致雌性果蝇个体的产生。*tra* 转录物的雄性特异性剪接会导致无功能活性蛋白质的产生，从而使 *dsx* pre-mRNA 按默认的雄性特异性剪接模式进行剪接，最终导致雄性果蝇个体的产生。Tra 和 Tra-2 连同一种或多种 SR 蛋白能够确保 *dsx* pre-mRNA 在雌性特异性剪接位点处发生剪接。这种定向性可能是绝大多数但不是全部选择性剪接的基础。

剪接的调控

通过上述两个生物系统的介绍我们已了解，同一条 pre-mRNA 分子可通过选择性剪接产生两种截然不同的产物。但选择性剪接并不只是个别事件，估计人类有超过一半的基因会经历选择性剪接，其中许多基因有两种以上的剪接方式，有的甚至达上千种。

图 14.41 举例说明了几种不同的选择性剪接模式。第一，转录物可以在不同的启动子处起始合成。在本例中，从第一个启动子处起始的转录可合成包含第一个外显子（A）的转录物。而从第二个启动子处起始的转录，其合成的转录物就没有第一个外显子（A）。第二，在转录过程中有些外显子（如外显子 C）会被遗漏，产生缺失该外显子的 mRNA 产物。第三，5′-剪接位点的可变性可导致某个外显子的部分序列（在本事例中为 D′）会在 mRNA 分子出现或缺失。第四，3′-剪接位点的可变性会导致某个外显子的部分序列（在本事例中为 F）会在 mRNA 分子出现或缺失。第五，内含子保留（retained intron）。如果内含子未被识别，那么该内含子就会留存在 mRNA 分子中，见图 14.41 下方的剪接模式。第六，多聚腺苷化作用会引起 pre-mRNA 的切割，导致下游外显子的缺失。例如，在 poly（A）位点 1 处进行切割会导致外显子 H 的缺失。有关 mRNA 的多聚腺苷化加工过程的内容将在第 15 章介绍。由此可见，如果在 6 个位点发生 2 个不同事件，则将会产生 $2^6 = 64$ 种结果。

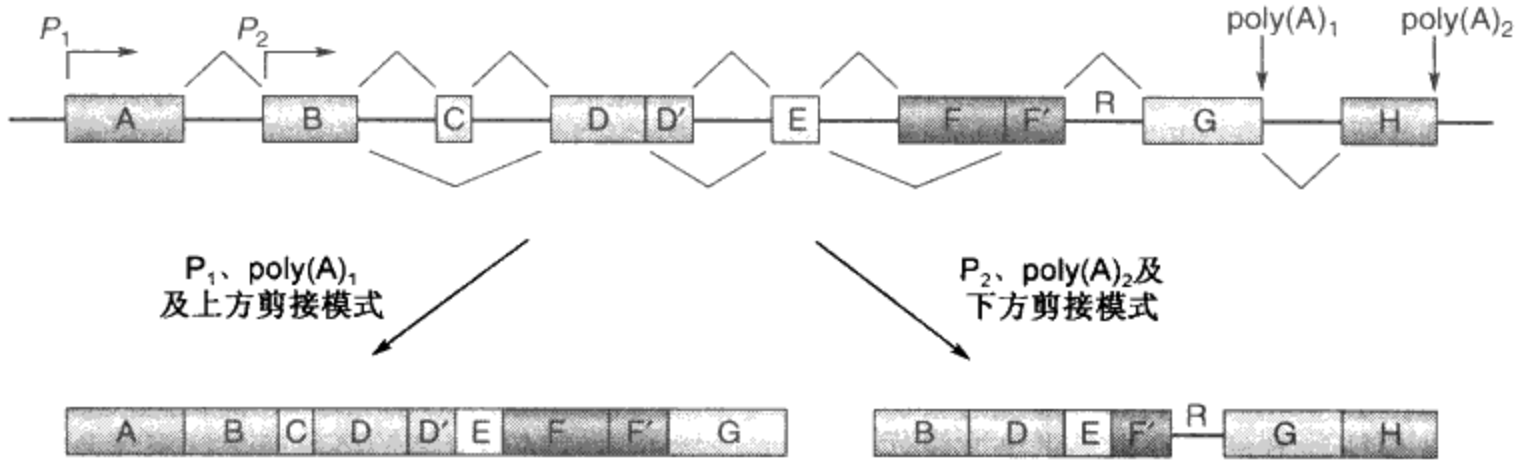


图 14.41 具有选择性启动子和多聚腺苷化位点的选择性剪接模式。图中只给出了可能会产生 64 种 mRNA 的可变剪接模式，6 个不同的决定性位点从左至右依此如下。① 若从第一个启动子处起始转录，则外显子 1能够出现在转录本中，若从第二个启动子处起始转录，则转录本会缺失外显子 1。② 外显子 C 识别的失败会导致下方所示剪接模式中对其剪接的省略。③ 对外显子 D 内 5′-选择性剪接位点（位于 D 和 D′之间）的识别会导致 D′在经历下方所示剪接模式后缺失。④ 对外显子 F 内 3′-选择性剪接位点（位于 F 和 F′之间）的识别会导致 F 在经历下方所示剪接模式后缺失。⑤ 对保留内含子（R）识别的失败会导致 R 在经历下方所示剪接模式后，留存在 mRNA 分子中。⑥ 多聚腺苷化作用会引起 pre-mRNA 在 poly（A）位点 1 处被切割，在经历上方所示剪接模式后，外显子 H 缺失。

选择性剪接显然会受到细胞的精细调控。否则，在雄性果蝇个体中 *dsx* pre-mRNA 可能就会发生雌性特异性剪接。这一切都暗示，在一种情况下被认为是外显子的序列，在另一种情况下则会被认为是

内含子的一部分。

那么，是什么机制能够促进这些信号在特定环境下被识别，而在另一种环境中又抑制对这些信号的识别？一部分答案是，正如我们刚才所了解到的，剪接因子促进了特定剪接位点的定向；另一部分答案是，外显子含有促进剪接的**外显子剪接增强子**（exonic splicing enhancer，**ESE**）和抑制剪接的**外显子剪接沉默子**（exonic splicing silencer，**ESS**）（内含子也存在着内含子剪接增强子和内含子剪接沉默子）。这些序列可能与细胞类型特异性或细胞发育阶段特异性蛋白质因子结合，或者与响应外部试剂（如激素）的蛋白质因子结合，这些结合作用可能会促进或抑制邻近位点的剪接反应。

果蝇的性别决定基因 *dsx* 为研究外显子剪接增强子的作用机制提供了一个很好的事例。该基因第 4 个外显子（图 14.39）的 3′-剪接位点信号不强，剪接因子 U2AF 难以识别。因此，在雄性果蝇中，外显子 4 因不能被识别而在成熟 mRNA 中缺失。但是在雌性果蝇中，*tra* 基因的蛋白质产物 Tra 以及 2 个 SR 蛋白能够与外显子 4 的 ESE 结合，并可能通过吸引剪接因子 U2AF 而激活对外显子 4 前面的 3′-剪接位点的识别。这样，外显子 4 就存在于成熟 mRNA 分子内。

目前，已有许多 ESE 被鉴定。发现 ESE 的途径之一就是进行序列敲除，然后观察特定位点的剪接缺失情况。鉴定 ESE 的另一个途径是利用功能性 SELEX 技术（第 5 章），该技术的基础是依赖于 ESE 促进剪接的能力，而不是与特殊分子结合的能力。Adrian Krainer 及同事克隆一段由外显子-内含子-外显子构成的 DNA 片段，其中第二个外显子携带 ESE 序列，研究者用一套 20 聚体的随机引物通过 PCR 取代 ESE，对数量达 1.2×10^{10} 的 DNA 片段进行转录，然后借助于凝胶电泳，从这些转录物中筛选能够在无细胞提取物中发生剪接的 RNA 分子。通过凝胶电泳可以将剪接和未被剪接的 RNA 区分开来。

功能性 SELEX 程序的缺点是，由于事先已知无细胞提取物中所加的 SR 蛋白种类，导致与未知蛋白作用的 ESE 被忽略。运用计算法可以解决这个问题：比对真正外显子和假外显子序列，找出在真正外显子中高频出现的短序列（6～10nt）。当然，ESE 不太可能出现在假外显子中，因为假外显子不需要剪接。但 ESE 一定会存在于需要剪接的真正外显子中（相反，ESS 倾向存在于假外显子中）。一旦假定的 ESE 被鉴定出来后，可以将 ESE 序列置于模式剪接底物中经常被跳过的外显子内，直接分析 ESE 促进剪接的能力。

ESE 易与 SR 蛋白相互作用，而 ESS 可与 hnRNP 相互作用，hnRNP 是结合有 hnRNA 的蛋白质。**hn RNP A1** 是与 ESE 活性有关的一种 hnRNP。分子生物学家已发现 hn RNP A1 至少有 3 种不同的作用机制（图 14.42），而且这 3 种机制都可发挥作用，不同的机制用于抑制不同外显子的剪接。

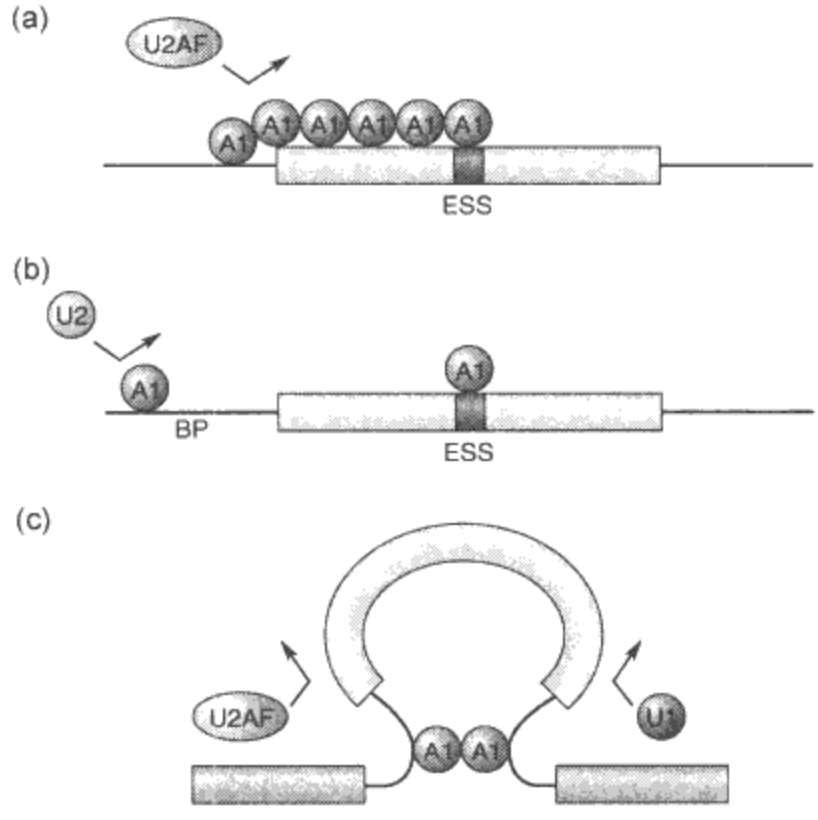


图 14.42　hnRNP A1 沉默剪接的模型。（a）A1 首先与 ESS 结合，协同结合的 A1 使其结合区域不断扩展，图中显示的是结合区向前一个内含子末端的 3′-剪接位点扩展，从而阻止 U2AF 的结合。（b）A1 与邻近内含子分支点（BP）附近的内含子沉默元件结合，阻止 U2 的结合。（c）A1 与黄色外显子侧翼内含子的内含子沉默元件结合，通过 A1 分子间的相互作用产生一个可隔离外显子 RNA 环，避免了外显子与剪接机构的结合。

第一个机制涉及一个 ESS：与外显子内 ESS 结合的 A1 能够引起其他 A1 的协同结合。这样，A1 就结合在整个外显子区域，覆盖了剪接信号，使之不能被剪接结构发现。另外两个机制涉及 A1 与内含子**沉默元件**（intronic silencing element）的结合。HIV *tat* 基因第三个外显子的剪接就是第二个机制的例证，A1 与内含子分支点附近的结合位点结合，阻止 U2 snRNP 在该位点的结合，导致剪接失败。在第三种机制中，A1 与外显子侧翼的两个内含子位点结合，通过两个 A1 分子间的相互作用使外显子游离于 RNA 环之外，使其被剪接机构所忽略。

如何鉴定 ESS？一种途径是借助计算方法，通过与真外显子的序列比对，寻找高频出现在假外显子中的序列。另一种途径就是直接寻找抑制剪接的

序列。Christopher Burge 及同事为鉴定 ESS 序列，利用绿色荧光蛋白（green fluorescent protein，GFP）基因的两个外显子构建了一个报告重组质粒（图 14.43）。在两个外显子中间存在着第三个外显子，如果该外显子出现在成熟 mRNA 分子中，GFP mRNA 就会被其间断，因而也就不能产生 GFP。根据上述推理，Burge 及同事在中间的外显子内引入一段 10bp 的随机序列，用重组质粒转化细胞，在荧光下观察细胞是否表达 GFP。

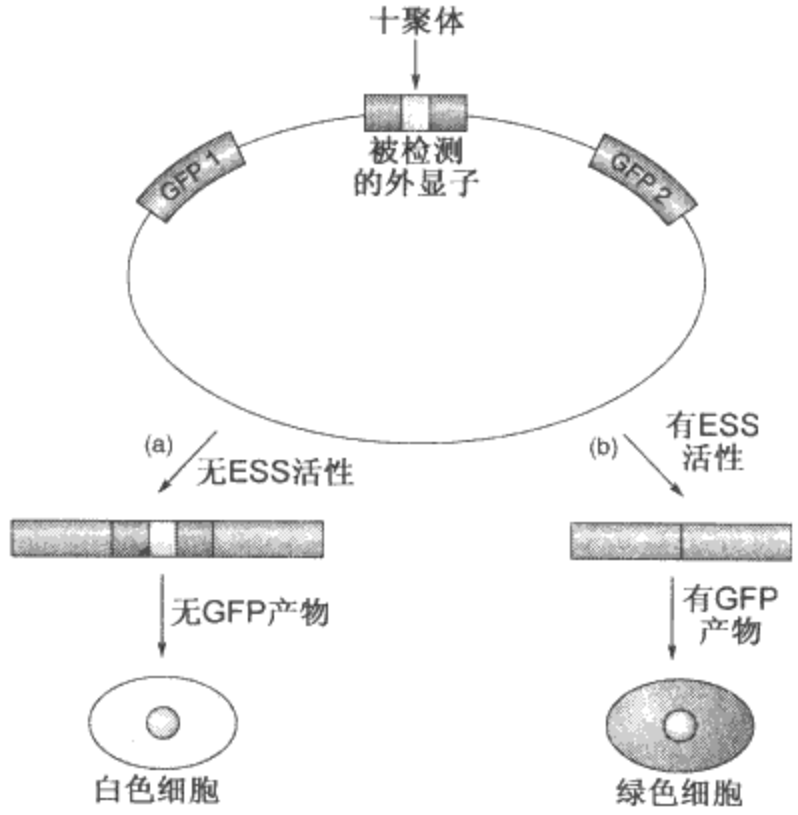


图 14.43　检测 ESS 活性的报告基因结构。Christopher Burge 及同事利用 GFP 基因的两个外显子构建了重组质粒，这两个外显子被中间内含子所隔离，该内含子含有测试外显子（红色），其中插入随机的 10bp 序列（黄色）。用重组质粒转化细胞，筛选表达 GFP 的细胞。(a) 如果 10bp 序列无 ESS 活性，测试外显子就会经历正常的剪接过程，出现在 GFP mRNA 分子中，且位于正常两个外显子的中间，导致 GFP mRNA 分子失活，不会产生功能性 GFP，因此细胞无荧光。(b) 如果 10bp 体序列有 ESS 活性，测试外显子就不能被剪接机构所识别，会连同内含子一起被切除，这样就会形成正常的 GFP mRNA 分子，细胞因表达 GFP 而呈现绿色。

绿色细胞表明该细胞能够表达 GFP，说明中间的外显子被切除，转录物经过剪接形成功能正常的 GFP mRNA 分子，进而表明中间外显子内的 10bp 序列在细胞内表现 ESS 活性。通过这种方法，Christopher Burge 及同事鉴定出了 141 个具有 ESS 活性的 10bp 序列，其中 133 个为单一序列。

小结　在高等真核生物中，选择性剪接是一种非常普遍的现象，选择性剪接可使同一个基因产生一种以上的蛋白质产物，是细胞调控基因表达的重要方式之一。这种调控基因表达机制的功能发挥需要剪接因子与剪接位点和分支点的结合、蛋白质与外显子剪接增强子（ESE）及外显子沉默子（ESS）间的相互作用。SR 蛋白倾向与 ESE 结合；hnRNP A1 倾向与 ESS 及内含子沉默元件结合。

14.3　RNA 的自我剪接

20 世纪 80 年代，分子生物最令人震惊的发现之一是，有些 RNA 分子在没有剪接体或其他蛋白质的帮助下能够自剪接。Thomas Cech（发音“Check”）及同事在研究纤毛原生动物四膜虫（*Tetrahytnena*）的 26S rRNA 基因时发现了这一现象。该基因与众不同的是它有一个内含子，但真正引人关注的是 1982 年的报道，即纯化的 26S rRNA 前体在体外能自剪接。事实上，这只是含有 **I 类内含子**（group I intron）的 RNA 自我剪接的第一个事例。在随后的研究中又发现了另一类含有 **II 类内含子**（group II intron）的 RNA，其中部分成员能够自剪接。

I 类内含子

为获得能进行自剪接的 RNA 分子，Cech 及同事克隆了包含内含子的 26S rRNA 基因的部分序列，构建了 2 个不同的重组质粒。利用 *E. coli* RNA 聚合酶进行质粒的体外转录，然后对转录的标记产物进行凝胶电泳（图 14.44），研究者观察到 4 个较大的 RNA 产物和 3 个较小的 RNA 产物，这 3 类较小的 RNA 产物大小与以前发现的线性和环形内含子（包括缺失 15nt 的线性内含子）的大小相对应。这个结果暗示 RNA 经历了剪接，剪除的内含子发生了环化。

这种剪接反应是由 RNA 自身催化进行的？还是 RNA 聚合酶也以某种方式参与了剪接？为回答这个问题，Cech 及同事在抑制剪接的多胺（精胺、亚精胺、腐胺）存在条件下进行 RNA 聚合酶反应。然后电泳产物，切取所有 4 个 RNA 条带以及存留在加样孔内的物质，纯化这些 RNA 产物，并在利于剪接的条件下（无多胺）孵育，再次电泳。对凝胶进行放射自显影处理后可见，条带 b ～d 中的 RNA 在孵育后产生了内含子，而原样品及条带 a 中的分子在孵育后无内含子产生（图 14.45）。所以，条带 b～d 显然是 26S rRNA 前体，在没有任何蛋白质甚至 RNA 聚合酶的参与下能够自剪接。

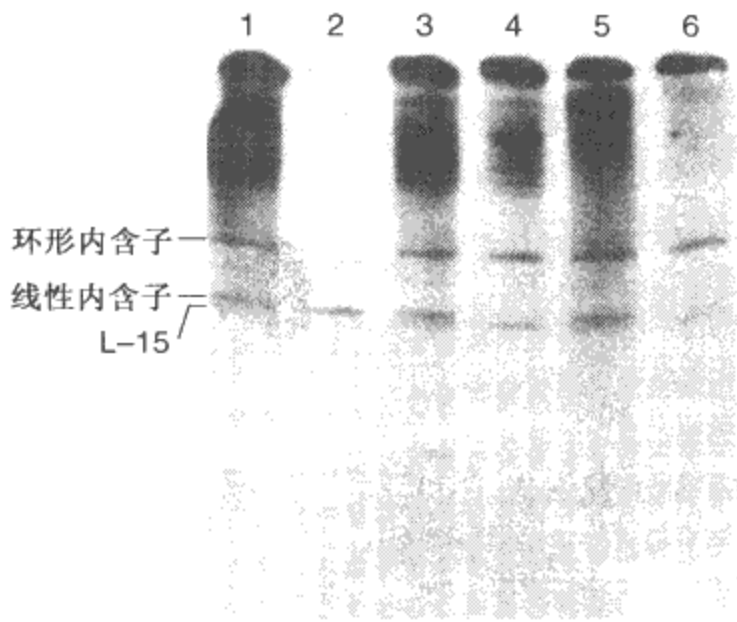


图 14.44　整合有 26S rRNA 内含子的质粒在体外转录后产生游离的内含子。Cech 及同事利用 *E. coli* 的 RNA 聚合酶，体外转录含有 rRNA 内含子的两种不同质粒，反应混合物中含有标记的核苷酸，用以标记转录产物。电泳标记 RNA 产物，放射自显影观察。泳道 1：*Eco*R I 切割的 pIVS13 的转录产物；泳道 2：线性及环形内含子标记；泳道 3：*Eco*R I 切割的 pIVS11 的转录产物；泳道 4：将泳道 3 的转录物在利于内含子环化的条件下（高温度及较高的 $MgCl_2$、硫酸铵浓度）孵育 30min；泳道 5：超螺旋 pIVS11 的转录物；泳道 6：将泳道 5 的转录物进行与泳道 4 相同的处理。泳道 5 中的大分子 RNA 难以分辨，是因为模板是非线性的，无确定的转录终点。图左侧标出了环形内含子、线性内含子以及在 5′端缺失 15nt 的线性内含子（L-15）的泳动位置。（*Source*：Kruger，K.，P. J. Grabowski，A. J. Zaug，J. Sands，D. E. Gottschling，and T. R. Cech，Self-splicing RNA：Autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* 31（Nov 1982）f. 2，p. 149. Reprinted by permission of Elsevier Science.）

我们称之为内含子的条带其大小是正确的，但是否真是我们所认为的内含子？Cech 及同事对该 RNA 的前 39 个核苷酸进行了测序，结果与内含子的前 39nt 完全一致。因此，这个 RNA 显然就是内含子。

Cech 研究小组还发现，线性内含子（我们一直在讨论的 RNA）可以自我环化。在提高温度、Mg^{2+} 及盐浓度条件下，至少有些纯化的线性内含子会发生环化。

到目前为止，我们已经了解到 rRNA 前体能够自我切除内含子，但它能否在切除内含子的同时将外显子连接在一起？Cech 及同事利用模式剪接反应实验证明，rRNA 前体在自我切除内含子的同时确实能将外显子连接到在一起（图 14.46a）。研究者克隆了四膜虫 26S rRNA 基因的部分序列，包含 303bp 的第一个外显子、全长内含子、624bp 的第二个外显子。将此基因片段连接到启动子能够被噬菌体 SP6 RNA 聚合酶所识别的载体中。在 [α-^{32}P] ATP 存在的条件下，经噬菌体 SP6 RNA 聚合酶作用于体外合成标记剪接前体。在剪接条件下，将 RNA 剪接底物分别在 GTP 存在和不存在两种情况下孵育，电泳产物。泳道 1 显示有 GTP 时形成的产物，存在有线性内含子和少量环形内含子。此外，出现了由拼接外显子形成的显著条带。相反，泳道 2 显示无 GTP 时，这些产物都不会形成，只有底物的存在。这正是所预期的结果，因为 I 类内含子的剪接是一个依赖 GTP 的反应，也充分证实了这些产物都是剪接反应的结果。总之，上述研究结果充分证实了剪接反应的发生，包括外显子的连接。

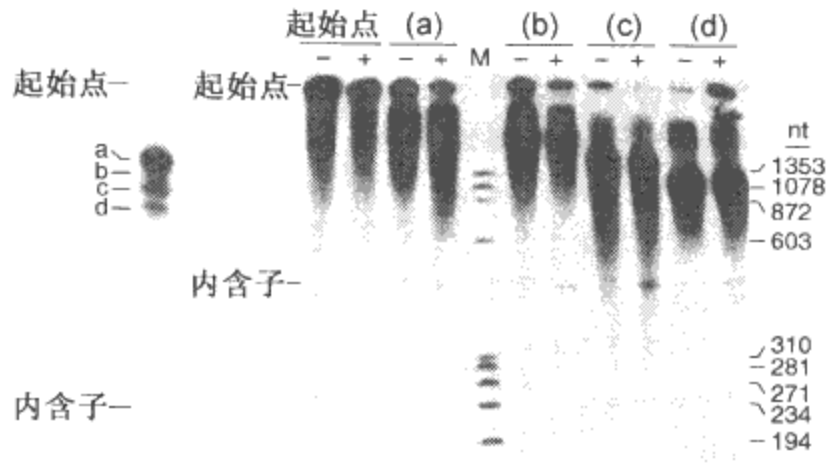


图 14.45　转录反应与剪接反应可区分开。（a）转录。Cech 及同事在抑制剪接的条件下（存在多胺）转录质粒 pIVS11，然后电泳标记的产物。出现了 4 个条带（在左侧标注为 a～d）。在该实验中 a 和 b 较难区分开。（b）剪接。Cech 及同事分别纯化图（a）中的起始点、a、b、c、d 5 个条带中的 RNA，一部分直接电泳（泳道－）；另一部分在利于剪接的条件下（无多胺）孵育，然后电泳（泳道＋）。RNA 的性质（起始点及条带 a～d）在图上方标出。起始点内的原样品及条带 a 中的分子在孵育后无内含子产生；而条带 b～d 的 RNA 在温育后则产生了内含子，但只是在利于剪接的条件下。（*Source*：Kruger K.，P. J. Grabowski，A. J. Zaug，J. Sands，D. E. Gottschling，and T. R. Cech，Self-splicing RNA：Autoexcision and autocylization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* 31（Nov 1982）f. 3，p. 150. Reprinted by permission of Elsevier Science.）

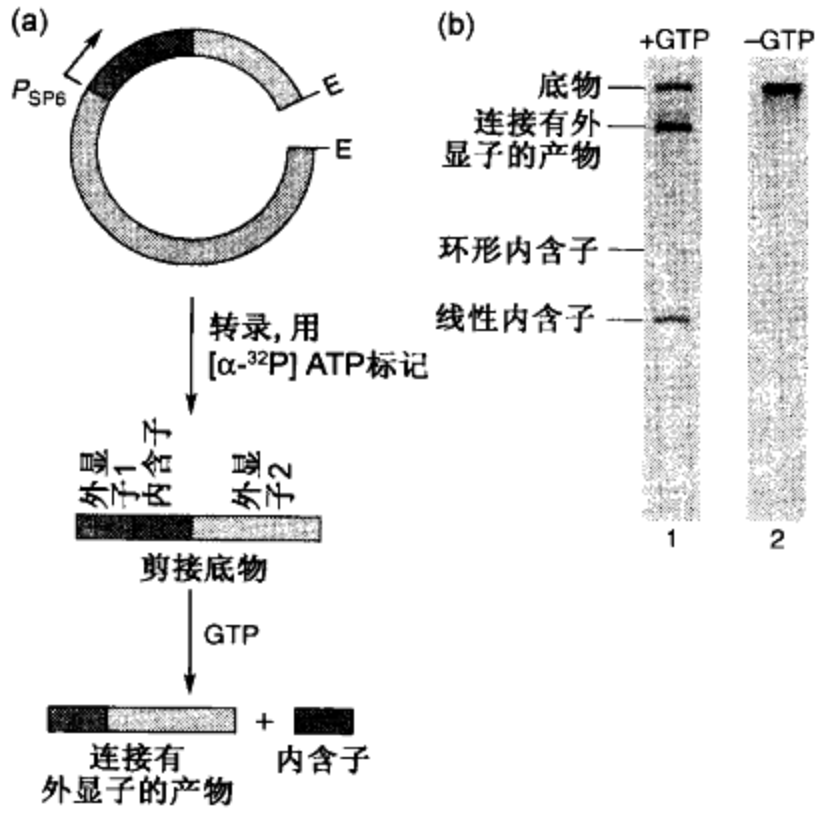


图 14.46　外显子连接的证明。（a）实验方案。Cech 及同事构建了包含四膜虫 26S rRNA 基因部分序列的重组质粒：长 303bp 的外显子 1（蓝色）、长 413bp 的内含子（红色）、长 624bp 的外显子 2（黄色）。用 *Eco*RI 酶切，线性化质粒，产生 *Eco*RI 酶切末端（E）。在［α^{32}P］ATP 存在的条件下，用噬菌体 SP6 RNA 聚合酶在体外转录线性质粒，合成标记的剪接底物。在剪接条件下，将剪接底物分别在 GTP 存在和不存在条件下孵育，然后电泳剪接产物，放射自显影鉴定标记的 RNA。（b）实验结果。在 GTP 存在时（泳道 1），除出现分别代表线性内含子和环形内含子的条带外，还出现了一条代表拼接外显子的显著电泳条带；在缺少 GTP 时（泳道 2），只出现了代表剪接底物的电泳条带。因此，外显子拼接可能是该 RNA 催化自剪接反应的一部分。(*Source*: (*b*) Inane, T., F. X. Sullivan, and T. R. Cech, Intermolecular exon ligation of the rRNA precursor of *Tetrahymena*: Oligonucleotides can function as 5′-exons. *Cell* 43(Dec 1985) f. 1a. p. 432. Reprinted by permission by Elsevier Science.)

Cech 小组的研究结果显示，26S rRNA 前体的剪接涉及在内含子的 5′端添加一个鸟嘌呤核苷酸。为证实缺少蛋白质参与的自我剪接也采用相同的机制，研究者进行了一个两部分实验（two-part experiment）研究。在第一部分实验中，它们于剪接和非剪接条件下，将剪接前体与［α^{32}P］GTP 共同孵育，然后电泳产物凝胶分析内含子是否被标记。图 14.47a 显示内含子确实被标记了，用［γ^{32}P］GTP 所做的相似实验其结果相同。在第二部分实验中，以相同方法用［α^{32}P］GTP 标记内含子的 5′端，然后对产物测序分析。结果发现，产物的核苷酸序列与线性内含子的核苷酸序列完全一致，只是在 5′端多出一个额外的 G（图 14.47b）。这个额外的 G 可被 RNase T1 切除，表明 G 是通过正常的 5′-3′-磷酸二酯键连接在内含子末端的。图 14.48 描述了四膜虫 26S rRNA 前体的剪接模型，示意了外显子的连接、线性内含子的形成过程。

我们已了解到，切除的内含子可自我环化。Cech 及同事研究发现，内含子的环化涉及线性内含子 5′端一段 15nt 片段的缺失。对此有 3 个证据：①内含子标记的 5′端没有出现在环形内含子中；②用RNase T1 处理线性内含子时，在其 5′端至少会产生 2 种（实际是 3 种）产物，而用 RNase T1 处理环形内含子时，则无这些产物的形成；③随着内含子的环化，包括 RNase T1 产物在内的 15 聚体 RNA 产物不断积累。

但环化并不是剪接的最后环节。环化之后，环化内含子又在成环时形成的磷酸二酯键处打开，然后切除 5′端约 4nt 小片段后再次环化。最后，内含子在刚形成的磷酸二酯键处再次打开，形成缩短了的线性内含子。

图 14.49 描述了切除的内含子环化与再次线性化的具体机制。注意，在 26S rRNA 前体的自我剪接过程中，在破坏一个磷酸二酯键的同时会形成一个新的磷酸二酯键，每一步自由能的改变几乎是零，不需要外界提供能量（如 ATP）。剪接过程的另一个特征是，环化分子转化为线性分子时破坏的磷酸二酯键恰是环化时形成的磷酸二酯键，表明这些磷酸二酯键具有特殊之处。RNA 分子的三维构型一定会使这些磷酸二酯键产生张力，从而使其在再次线性化时易于被打开，这种张力有助于解释 RNA 的催化能力。

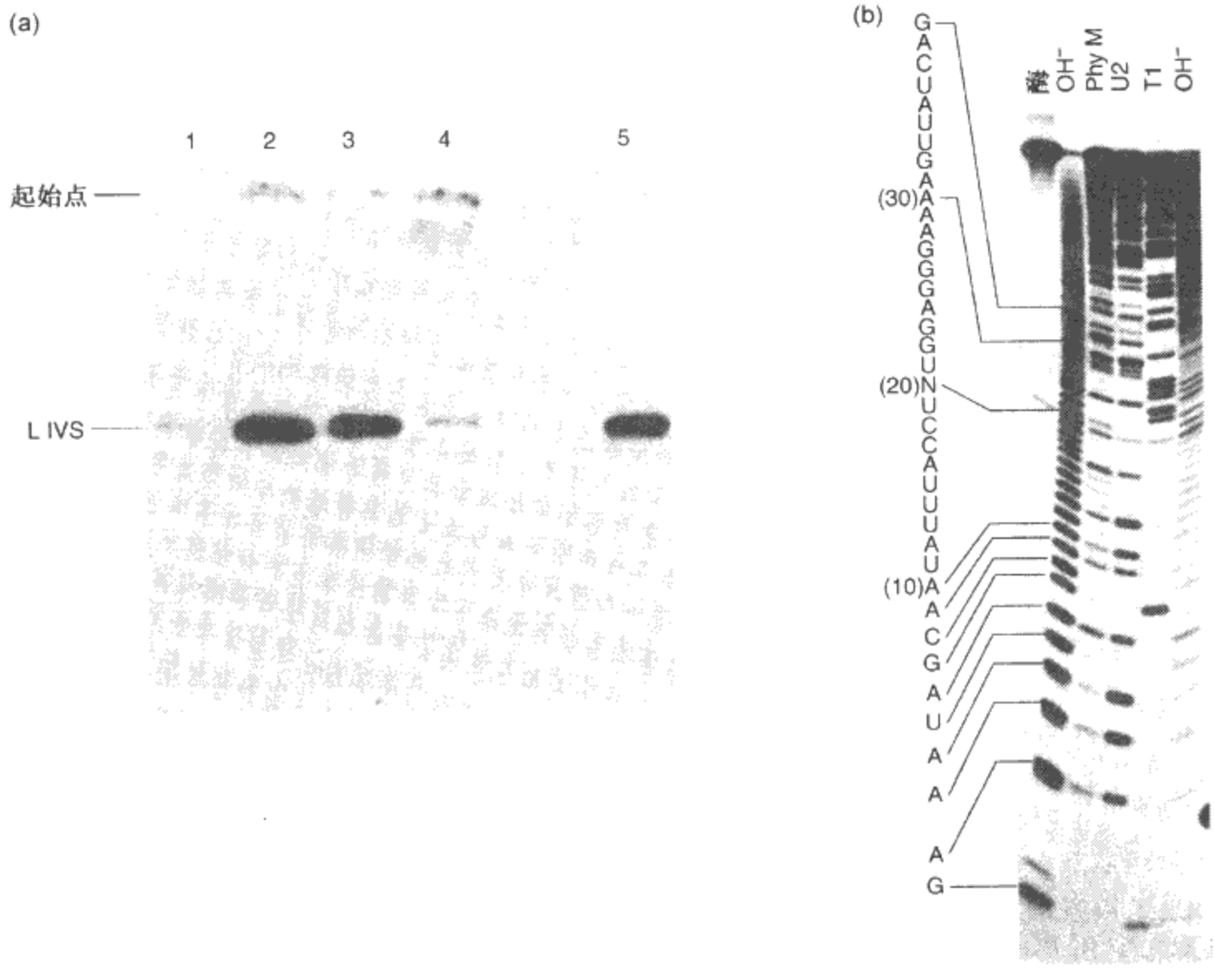


图 14.47　GMP 连于被切下内含子的 5′端。（a）在剪接过程中，用放射性 GTP 标记内含子。在不能发生剪接反应的条件下，Cech 及同事以非标记的核苷酸为底物，体外转录质粒 pIVS11。分离非标记的 26S rRNA 前体，在允许发生剪接反应的条件下，加入［α-^{32}P］GTP 后保温孵育。用 Sephadex G-50 层析分离产物，将分离的片段进行电泳并放射自显影处理。泳道 1～4：从 Sephadex G-50 分离中获得的连续片段；泳道 5：线性内含子的分子标记；泳道 2～3：含有大量的线性内含子，因其被标记，说明线性内含子结合有放射性的 G。（b）标记内含子的测序。Cech 及同事利用酶法对该 RNA 的 5′端测序。在强碱（OH^-）条件下降解 RNA，使之分解为单个核苷酸；RNase Phy M 切割 A 和 U；RNase U2 切割 A；RNase T1 切割 G。对 RNA 样品的处理方法在图上方标出，左侧为推导的序列，5′-G 位于底部。（*Source*：Kruger K.，P. J. Grabowski，A. J. Zaug，J. Sands. D. E. Gottschling and T. R. Cech. Self-splicing RNA：Autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* 31（Nov 1982）f. 4，p. 151. Reprinted by permission of Elsevier Science.）

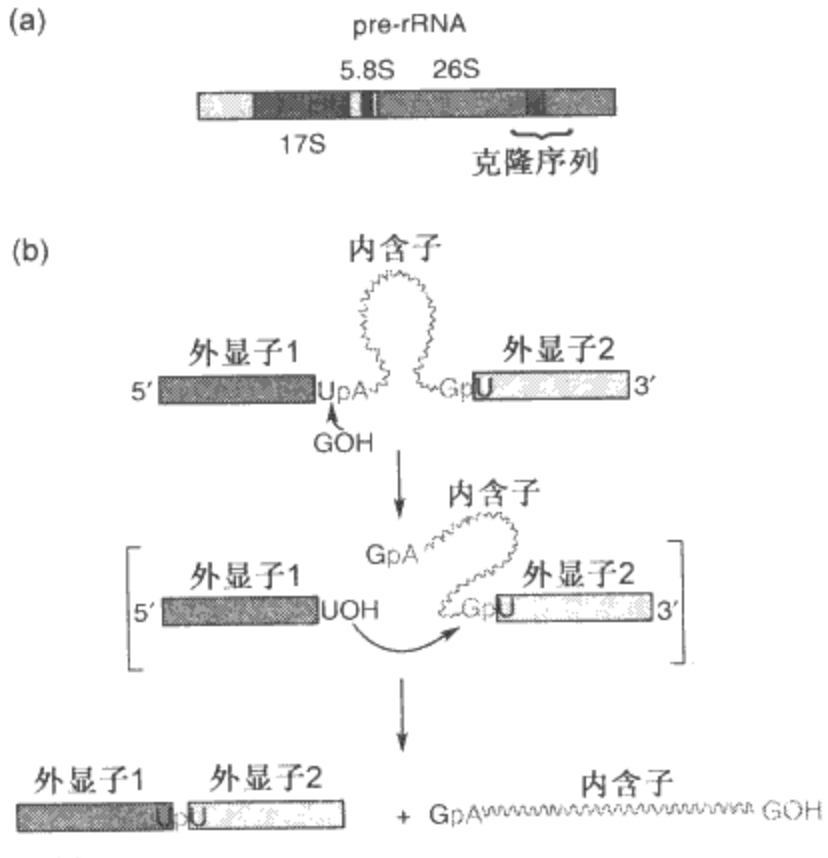


图 14.48　四膜虫 rRNA 前体的自剪接。（a）rRNA 前体的结构。包括 17S、5.8S 和 26S 序列，内含子（红色）存在于 26S 区域。方括号标出用于实验的克隆片段。（b）自剪接示意图。第一步（上方）：鸟嘌呤核苷酸攻击内含子 5′端的腺嘌呤核苷酸，释放外显子 1（蓝色），产生假定的中间体。第二步：外显子 1 的 3′-OH 攻击外显子 2（黄色），释放线性内含子（红色），两个外显子连接在一起。最后，在一系列反应中（未显示）线性内含子在 5′端丢失 19nt。

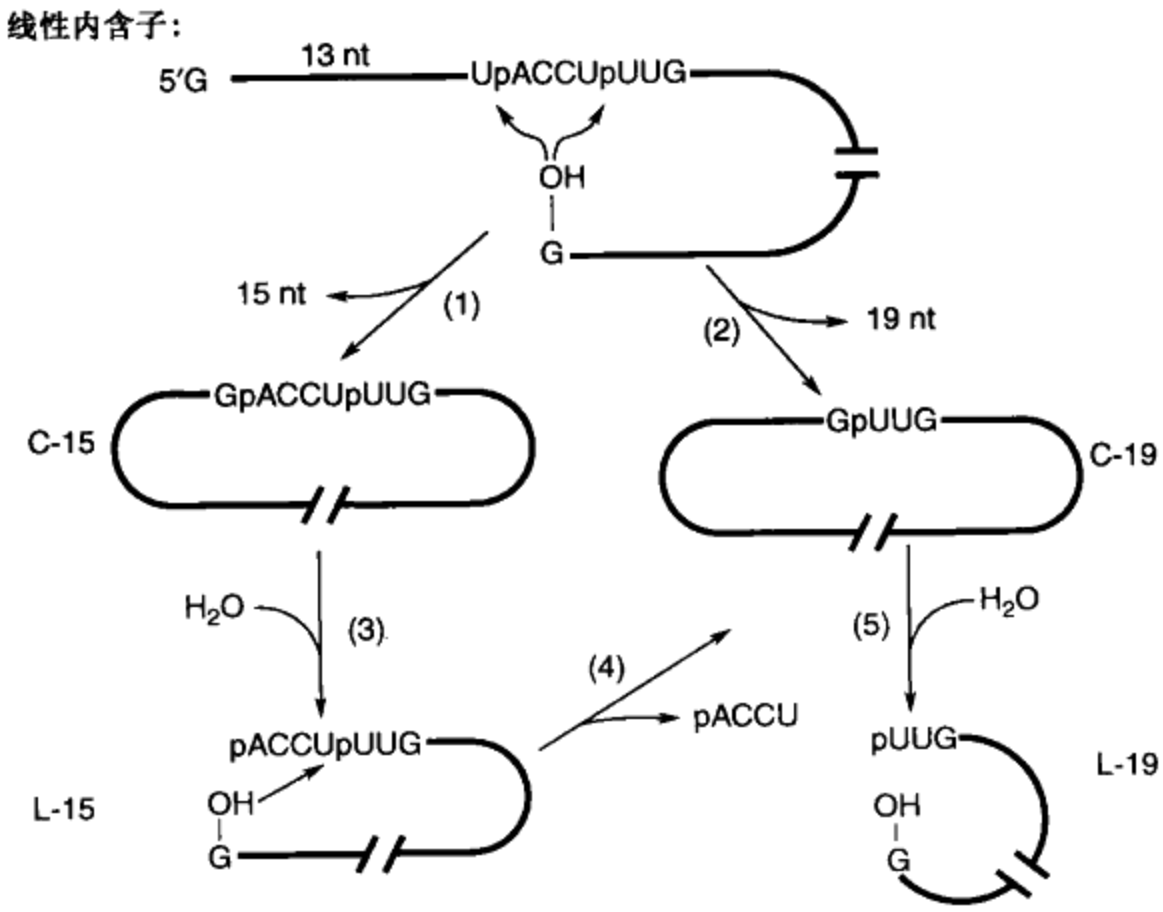
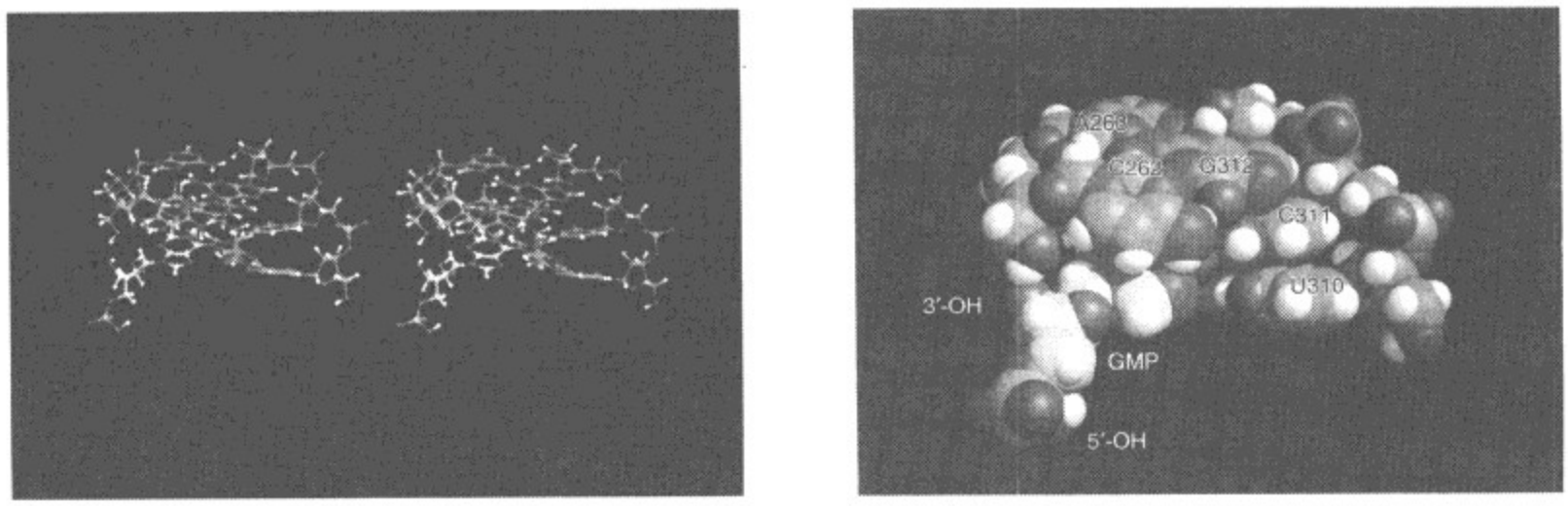


图 14.49　线性内含子的命运。从 26S rRNA 前体剪切下的线性内含子可通过两条途径发生环化。反应 1（绿色箭头）：3′端的鸟嘌呤攻击 U-15 和 A-16 之间的磷酸二酯键，切下长 15nt 的片段，产生环形内含子（C-15）；反应 2（蓝色箭头）：末端 G 再向前 4nt 处多攻击内含子 4nt，切下 19nt 片段，并留下较小的环形内含子（C-19）；反应 3：C-15 将在环化时形成的磷酸二酯键打开，产生线性内含子（L-15）；反应 4：L-15 的末端 G 攻击 5′端前两个 U 之间的磷酸二酯键，产生环形内含子 C-19；反应 5：C-19 解环而成为线性内含子 L-19。

初看起来，剪接体内含子与 I 类内含子的剪接机制存在显著的差别：在剪接的第一步，I 类内含子使用外源鸟嘌呤核苷酸，而剪接体内含子使用存在于内含子自身序列中的鸟嘌呤核苷酸。然而，当我们仔细研究这两种剪接机制时，发现它们之间的差异并不是很大。Michael Yams 及同事用分子模拟技术（molecular modeling technique）预测了四膜虫 26S rRNA 与 GMP 结合时的最低能量构象。研究者指出，内含子部分序列折叠成双螺旋结构，其口袋通过氢键与鸟嘌呤核苷酸结合（图 14.50）。鸟苷酸与内含子结合十分紧密，其作用的本质与剪接体内含子腺嘌呤的作用相同。当然，由于鸟苷酸与内含子之间无共价键，因此不会形成套索状分子。

在 RNA 分子自剪接现象发现以前，生物化学家们一直认为只有蛋白质才具有酶的催化功能。Sidney Altman 在几年前就发现，能在 tRNA 前体 5′端切除额外核苷酸的 **RNase P** 包含 RNA 组分 **Ml**，但 RNase P 也存在蛋白质组分，或许此蛋白质组分具有酶的催化活性。1983 年，Altman 通过证实发现 Ml RNA 就是 RNase P 的催化组分（第 16 章）。RNase P 以及能够自我剪接的 RNA 是具有催化活性的 RNA，因此称之为核酶（ribozyme）。事实上，从严格意义上讲，I 类内含子参与的剪接反应并不是酶促反应，因为在反应结束后，RNA 自身发生了改变。而真正的酶在反应结束后自身不会发生任何改变。四膜虫 26S rRNA 前体剪接后，形成的线性 I 类内含子可以像真正的酶一样发挥功能，将一个核苷酸添加到寡聚核苷酸上，并将其从寡聚核苷酸上切除。

小结　I 类内含子，如四膜虫 26S rRNA 前体中的内含子，能够在体外无其他蛋白质参与的情况下进行自我剪接。该剪接反应由鸟嘌呤核苷酸攻击内含子的 5′-剪接位点而引发，该 G 被添加到内含子的 5′端，并释放第一个外显子。在第二步反应中，第一个外显子攻击 3′-剪接位点，将两个外显子连接在一起，并释放线性内含子，该内含子发生两次环化，且每次环化时都会丢失部分核苷酸，环化的内含子最终转化为线性分子。



(a) (b)

图 14.50　位于 26S rRNA 内含子口袋内的 GMP 的两种视图。(a) 交叉立体图，交叉近视指导两图像融合在一起时，一段时间后即可观察到三维立体图。RNA 的碳原子用绿色表示；鸟嘌呤核苷酸的碳原子用黄色表示；磷原子用淡紫色表示；其他原子用标准色表示。 (b) 空间填充模型。颜色标注同图 (a)。(*Source*: Yarus, M., I. lllangesekare, and E. Christian, An axial binding site in the *Tetrahymena* precursor RNA. *Journal of Molecular Biology*. 222 (1991) f. 7c-d, p. 1005. by permission of Academic Press.)

II 类内含子

根据所包含的特定保守序列，真菌线粒体基因的内含子最初被划分为两类：I 类和 II 类。随后的研究发现，许多生物的线粒体基因和叶绿体基因含有 I 类内含子和 II 类内含子，含这两类内含子的部分 RNA 分子可以进行自剪接，但剪接机制不同。I 类内含子的自剪接反应由游离的鸟嘌呤核苷酸启动；II 类内含子的自剪接反应由内含子内部的腺苷酸残基启动，通过分子内攻击而形成套索结构。

II 类内含子的剪接反应与核 pre-mRNA 剪接体的剪接反应十分相似，均可形成套索结构，位于剪接复合体内的 mRNA 前体与 II 类内含子在整体结构上也高度相似，从图 14.22 所示结构可以看出。这暗示 snRNP 与 II 类内含子的催化活性部分在功能上的具有相似性，也反映了 RNA 之间具有共同的进化起源。事实上，已有研究者提出，细胞核 pre-mRNA 的内含子起源于细菌的 II 类内含子，携带内含子的细菌在侵染到真核细胞的祖先细胞后进化为真核细胞内的线粒体。当在古菌、蓝细菌及紫细菌中发现 II 类内含子后，这一假说受到了更大的关注。如果我们假定 II 类内含子的起源要早于这两个细菌类群的共同祖先的起源，那么，这些 II 类内含子一定存在于细菌内，而这些细菌就是现代真核细胞细胞器的祖先。然而，这种趋同进化的共有机制还只是一种可能性。

小结　II 类内含子的自我剪接需要通过 A 分支的套索中间体的介导，类似于剪接体的套索分子。剪接体系统和 II 类内含子的二级结构具有惊人的相似性。

总结

细胞核 mRNA 前体的剪接要通过套索状或分支状中间体介导。除了在内含子的 5′端和 3′端具有保守序列外，在分支点处也存在分支点保守序列。在酵母中，固有的保守序列为 UACUAAC。高等真核生物的分支点序列 YNCURAC 的保守性不强，但不管怎样，序列中的最后一个 A 为分支点核苷酸。酵母的分支点序列还能将其下游的 AG 确定为 3′-剪接位点。

RNA 的剪接是在剪接体中进行的，酵母剪接体和哺乳动物剪接体的沉降系数大体分别为 40S 和 60S。遗传学实验表明，U1 snRNA 与 mRNA 前体 5′-剪接位点的互补配对是进行正常剪接所必需的，但并不是剪接发生的充分条件。通过碱基互补配对，U6 snRNP 也能结合在内含子的 5′端，且两者的结合发生在套索状中间体形成之前，但剪接的第一步完成后，其性质可能发生了改变。U6 与剪接底物的结合是进行正常剪接反应所必需的，在剪接过程中，U6 也可与 U2 结合。

U2 snRNA 能够与剪接分支位点的保守序列进行碱基互补配对。这种碱基配对是剪接反应所必需的。U2 与 U6 形成的关键碱基对构成了螺旋 I 结构区，对参与剪接的 snRNP 具有定位功能。U4 snRNA 能够与 U6 进行碱基配对，其作用可能是与 U6 结合，直至 U6 需要参与到剪接反应时，U4 才与 U6 解离。U5 snRNP 能与一个外显子的最后一个碱基和下一个外显子的第一个碱基发生交联。可能的作用是将 5′-剪接位点和 3′-剪接位点排列在一起便于剪接。

即将进行第二步剪接反应的剪接体（底物、U2、U5 和 U6）的结构特征可以从处于相同剪接阶段的 II 类内含子的结构特点推知。因而，II 类内含子位于催化活性中心的元件似乎可以被剪接体的 snRNP 所取代，这些元件可能拥有剪接体的催化活性。

剪接体催化中心包括 Mg^{2+}、通过 U2 snRNA、U6 snRNA 和内含子分支点序列间碱基配对形成的复合体。这在三类无蛋白质组分的 RNA 片段可催化与第一步剪接的有关反应。

剪接体循环包括剪接体的组装、剪接活性和解体。组装起始于 U1 与剪接底物的结合形成定向复合体。U2 是第二个加入到复合体中的 snRNA，随后是其他 snRNA 的结合。其中 U2 的结合需要 ATP 供能。当 U6 与 U4 分离后，U6 便取代 U1 与 5′-剪接位点结合，这个 ATP 依赖的过程激活剪接体，并释放 U1 与 U4。参与剪接的 5 类 snRNP 均含有相同的 7 个 Sm 蛋白和一套自身特异蛋白。U1 snRNP 的结构显示，Sm 蛋白为环形结构，可以结合其他蛋白质。

一类少数具有 5′-剪接位点和分支点的内含子可在 U11、U12、U4 atac、U6 atac 等变体 snRNA 的帮助下完成剪接。剪接因子 Slu7 是选择正确 3′-剪接位点所必需的。当 Slu7 缺失时，在正确 3′-剪接位点 AG 处的剪接受到特异性抑制，而发生在距分支点 30nt 之内错误 AG 处的异常剪接则被激活。对 3′-剪接位点的识别还需要 U2AF 的参与，65kDa 的 U2AF 亚基结合在分支点下游多聚嘧啶碱基区，35kDa 的 U2AF 亚基与 3′-剪接位点的 AG 结合。

RNA 结合蛋白能够确保剪接底物在给定的位点处被剪接。这些 RNA 结合蛋白与剪接底物结合后，可能会招募其他剪接体组分的结合，首先是 U1 的结合。例如，SR 蛋白 SC35 和 SF2/ASF 分别定向人 β-球蛋白 pre-mRNA 和 HIV tat pre-mRNA 的剪接。

在酵母定向复合体中，分支点桥接蛋白（BBP）能够与内含子 5′端的 U1 snRNP 和邻近内含子 3′端的 Mud2p 发生结合。此外，BBP 还能与内含子 3′端附近的 RNA 序列结合。因此，BBP 能够将内含子的两个末端桥接在一起，在剪接之前将内含子界定出来。哺乳动物的 BBP 同源蛋白 SF1（mBBP）在哺乳动物定向复合体中具有相同的桥接作用。

RNA 聚合酶 II RpbI 亚基的 C 端功能域（CTD）能够促进外显子界定的底物剪接，但不会促进内含子界定的底物剪接。CTD 与剪接因子结合，从而将剪接因子装配到外显子的两个末端，引发剪接反应。

许多真核生物基因的转录物可进行选择性剪接，这对基因的蛋白质产物有着重要的影响。例如，通过选择性剪接可产生分泌型和膜结合型两种不同的蛋白质，甚至能够产生具有活性和无活性之分的蛋白质。在果蝇的性别决定途径中，3 个基因的转录产物就经历了选择性剪接。*tra* 基因转录物经雌性特异性剪接后产生功能活性 Tra 蛋白，引发 *dsx* pre-mRNA 的雌性特异性剪接，最终导致雌性果蝇个体的产生。*tra* 转录物的雄性特异性剪接会导致无功能活性蛋白质的产生，从而使 *dsx* pre-mRNA 按默认的雄性特异性剪接模式进行剪接，最终导致雄性果蝇个体的产生。Tra 和 Tra-2 连同一种或多种 SR 蛋白能够确保 *dsx* pre-mRNA 在雌性特异性剪接位点处发生剪接。这种定向性可能是绝大多数但不是全部选择性剪接的基础。

在高等真核生物中，选择性剪接是一种非常普遍的现象，选择性剪接可使同一个基因产生一种以上的蛋白质产物，是细胞调控基因表达的重要方式之一。这种调控基因表达机制的功能发挥需要剪接因子与剪接位点和分支点的结合、蛋白质与外显子剪接增强子（ESE）及外显子沉默子（ESS）间的相互作用。SR 蛋白质倾向与 ESE 结合；hnRNP A1 倾向与 ESS 及内含子沉默元件结合。

I 类内含子，如四膜虫 26S rRNA 前体中的内含子，能够在体外无其他蛋白质参与的情况下进行自我剪接。该剪接反应由鸟嘌呤核苷酸攻击内含子的 5′-剪接位点而引发，G 被添加到内含子的 5′端，并释放第一个外显子。在第二步反应中，第一个外显子攻击 3′-剪接位点，将两个外显子连接在一起，并释放线性内含子，该内含子发生两次环化，且每次环化时都会丢失部分核苷酸，环化的内含子最终转化为线性分子。

II 类内含子的自我剪接需要通过 A 分支的套索中间体的介导，类似于剪接体的套索分子。剪接体系统和 II 类内含子的二级结构具有惊人的相似性。

复习题

1. 叙述 R 环实验并给出结果，证明内含子也能被转录。

2. 图示说明剪接的套索机制。

3. 给出能够说明被切离的内含子是环形或套索状的凝胶电泳实验数据。

4. 给出凝胶电泳实验数据，区分套索状剪接中间体（内含子-外显子 2 形成的中间分子）和套索状产物（被切离的内含子）。

5. 套索模型推测套索状剪接中间体含有分支点核苷酸。叙述证实这一推测的实验并给出结果。

6. 叙述实验并给出结果，证明酵母内含子序列 UACUAAC 是剪接反应所必需的。

7. 叙述实验并给出结果，证明酵母内含子序列 UACUAAC 指定其下游 AG 为剪接位点。

8. 在套索剪接模型中，内含子序列 UACUAAC 发挥什么作用?

9. 叙述实验并给出结果，证明酵母剪接体沉降系数为 40S。

10. 叙述实验并给出结果，证明 U1 snRNA 与 5′外显子-内含子边界序列互补配对是剪接反应所必需的。

11. 叙述实验并给出结果，证明 U1 snRNA 与 5′外显子-内含子边界序列互补配对不是剪接反应的充分条件。

12. 为确保剪接反应的发生，除 U1 和 U5 外，还有哪些 sn RNP 必须与 5′外显子-内含子边界序列结合? 利用两个交联实验数据证实你的结论。

13. 在 RNA 交联实验中，使用 4-thioU 和补骨脂素的两个主要区别是什么?

14. 叙述实验并给出结果，证明 U2 snRNA 与分支点序列互补配对是剪接反应所必需的。在这个实验中，为什么细胞不能只有一个 U2 突变基因拷贝?

15. 除了能与 pre-mRNA 进行碱基互补配对外，U6 还能与哪两类 snRNA 互补配对?

16. 叙述实验并给出结果，证明在剪接过程中，U5 既能与上游外显子的 3′端结合，又能够与下游外显子的 5′端结合。实验方案需明确鉴定出参与该过程的 RNA 种类，不仅仅依靠电泳迁移率。

17. 叙述实验并给出结果，确定 U5 分子中能够与 pre-RNA 发生交联的碱基，并阐述实验结果。

18. 归纳总结有关在剪接体剪接反应中 Mg^{2+} 参与催化的证据。

19. 归纳总结有关含有剪接体 RNA 片段的混合物能够催化第一步剪接反应的证据。

20. 图示说明在第二步剪接反应开始前 pre-RNA 在剪接体中存在的特点，同时展示 pre-RNA 与 U2、U5、U6RNP 的相互作用，这种结构特征与哪一类自我剪接 RNA 的剪接中间阶段相似?

21. 叙述实验并给出结果，证明 U1 是第一个与剪接底物结合的 snRNP。

22. 叙述实验并给出结果，证明其他所有 snRNP 结合到剪接体都需要 U1 的协助，以及 U2 的结合需要 ATP。

23. 什么是 Sm 蛋白?

24. 少数剪接体的特征是如何显示 snRNA 与 pre-RNA 位点间碱基配对的重要性的?

25. 叙述实验并给出结果，证明对 3′-剪接位点中正确 AG 的选择需要 Slu7 参与。

26. 叙述筛选参与剪接定向过程的剪接因子的剪接定向分析实验，并阐述实验结果。

27. 叙述分析两个已知蛋白质间相互作用关系的酵母双杂交分析实验方案。

28. 叙述通过酵母双杂交系统寻找能与已知蛋白质发生相互作用的未知蛋白质的实验方案。

29. 叙述酵母双杂交实验并给出结果，证明酵母分支点结合蛋白（BBP）能够与其他两种蛋白质发生相互作用。这两种蛋白质是什么? 相对于内含子末端而言，这两个蛋白质结合在定向复合体的什么位置上?

30. 叙述实验并给出结果，证明 RNA 聚合酶 II CTD 能够促进外显子界定模式中的 pre-RNA 剪接。

31. 图示说明免疫球蛋白 μ 重链转录物的选择性剪接过程。主要体现只参与一种选择性剪接途径的外显子，无须体现参与各剪接途径的外显子。通过两种剪接途径产生的蛋白质有何不同?

32. 叙述用于鉴定具有外显子剪接沉默子功能的序列的计算方法及实验方法。

33. 叙述实验并给出结果，证明 I 类内含子能够自我剪接。

34. 叙述实验并给出结果，证明线性 I 类内含子能够自我环化。

35. 叙述实验并给出结果，证明含有 I 类内含子的 RNA 能进行自我环化。

36. 叙述实验并给出结果，证明鸟嘌呤核苷酸添加在切离内含子末端的实验方案，并阐述实验结果。

37. 图示说明含有 I 类内含子的 RNA 分子的自我剪接步骤，无须体现内含子的环化过程。

38. 图示说明四膜虫 26S pre-rRNA 从切下线性内含子起到 L-18 内含子形成的整个过程。不必体现 C-15 中间体。

分析题

1. 你正在研究的基因含有 1 个大的内含子和 2 个较短的外显子，在以下列分子为研究对象的情况下，阐述 R 环实验的结果：

a. mRNA 与单链 DNA；

b. mRNA 与双链 DNA；

c. mRNA 前体与单链 DNA；

d. mRNA 前体与双链 DNA。

2. 你发现一类新的内含子，其在剪接过程中不需要任何蛋白质的协助，但需要几类小 RNA 分子，其中之一是 V3，具有能够与 3′-剪接位点互补的 7nt 序列 CCUUGAG。你推测 V3 与 3′-剪接位点互补配对是剪接反应所必需的。设计能够证明你的假设的

实验方案，简要阐述确定的结果。

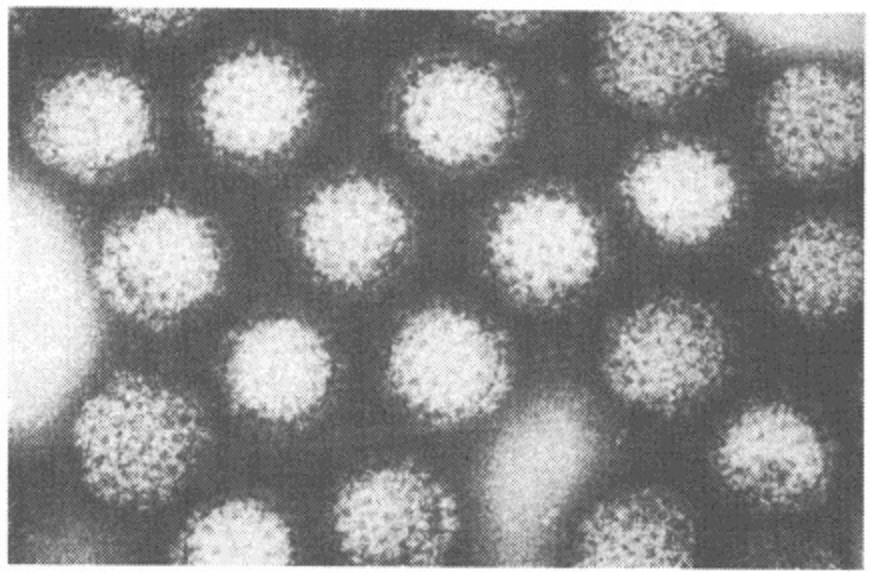
3. 图示说明 RNA 酶 T1（或 T2）的作用机制。RNA 分子在碱性条件下水解的机制与 RNA 酶 T1（或 T2）的作用机制相同，阐释为什么 DNA 在碱性条件下不能发生水碱?

4. 你正在研究人类 β-地中海贫血这种疾病，患者不能合成 β-球蛋白。你发现患者 β-球蛋白基因的编码区正常，但 mRNA 分子比正常情况长出约 100 多个核苷酸。对患者 β-球蛋白基因测序发现，在基因的第一个内含子中有一个碱基发生了改变。提出一个假说，阐释 β-地中海贫血患者缺失 β-球蛋白的原因。

5. 对于图 14.41 描述的基因结构而言，如果去除 P_2 和 poly（A）$_1$，只保留一个启动子（P_1）和一个多聚腺苷化位点［poly（A）$_2$］，该基因的转录本通过剪接可形成多少种 mRNA 分子?

翻译　徐启江　　校译　李玉花

第 15 章　mRNA 加工 II：加帽和多聚腺苷酸化



人类呼肠孤病毒的假彩色电镜图像（×1000 000）© *K. G. Murti/Visuals Unlimited*

除剪接外，真核细胞还要对 RNA 进行其他形式的加工，如 mRNA 要经历被称为加帽（capping）和多聚腺苷酸化（polyadenylation）的两个加工过程；在加帽过程中，一个特殊的封阻核苷酸（帽子）被加在了前体 mRNA（pre-mRNA）的 5′端，在多聚腺苷酸化过程中，一串 AMP［poly（A)］被加在前体 mRNA 的 3′端。这些过程对 mRNA 行使正常功能是至关重要的，同时也是这一章我们要讨论的内容。

15.1　加帽

1974 年，研究者发现在不同的真核生物以及病毒体内，其 mRNA 分子都存在甲基化现象。而且，大量的甲基化位点都集中在 mRNA 分子的 5′端，在结构上我们称之为**帽子**（cap）。这一节将介绍“帽子”的结构和形成过程。

帽子的结构

在基因克隆技术已发展成为实验室的常规技术之前，对病毒 mRNA 分子进行纯化和研究要比真核细胞容易得多。因此，第一个帽子序列是在病毒的 RNA 中被发现的。Bernard Moss 及同事用下面的方法分离得到了牛痘病毒的 mRNA 并发现了其中的帽子序列。他们用带有［^{3}H］的 S-腺苷甲硫氨酸（S-adenosylmethione，AdoMet，一种甲基供体）或者带有^{32}P 的核苷酸标记 RNA 甲基化基团，然后对这些经同位素标记的 RNA 进行水解。之后发现大部分水解产物都是单核苷酸，但帽子序列可以通过 DEAD-纤维素亲和层析加以分离。图 15.1 展示了当时对牛痘病毒 mRNA 的帽子序列进行色谱分析的结果。结果表明这些帽子序列带有大约 5 个负电荷。图 15.1b 中红色曲线是^{3}H（甲基化）标记的结果，蓝色曲线是^{32}P 标记的结果，二者的吻合表明帽子序列被甲基化了。Aaron Shatkin 小组在分析呼肠孤病毒的 RNA 帽子序列之后也得到了类似的结果。

为了确定呼肠孤病毒中帽子的结构，Yasuhiro Furuichi 和 Kin-Ichiro Miura 进行了如下实验。他们发现这个帽子序列能被［β，γ-^{32}P］ATP 标记却不能被［γ-^{32}P］ATP 标记。这一结果表明 β 位的磷酸基团可以被帽子序列保留，而 γ 位的却不行。在一个 RNA 分子中，只有第一个碱基仍然保留有 β 位磷酸基团，因此这一结果证实了帽子序列是在 RNA 的 5′端。但这个 β 位的磷酸基团显然是被一个未知的 X 物质所保护着，因为它不能被碱性磷酸酯酶（alkaline phosphatase）去除。

由此产生了下一个问题：X 物质是什么？研究发现这一保护基团可以被磷酸二酯酶（phosphodiesterase）移除，而磷酸二酯酶既可以消化磷酸二酯键也可以消化磷酸酐键（如核苷酸内 α 和 β 位的磷酸键）。经过磷酸二酯酶消化产生带有负电荷的类似 Xp 的化合物。之后，Furuichi 和 Miura 用单磷酸酯酶（phosphomonoesterase）去掉了 Xp 中的磷酸基团，只剩下 X，然后对这一化合物进行纸电泳，随后又进行纸色谱分析。图 15.2 显示 X 的电泳距离与

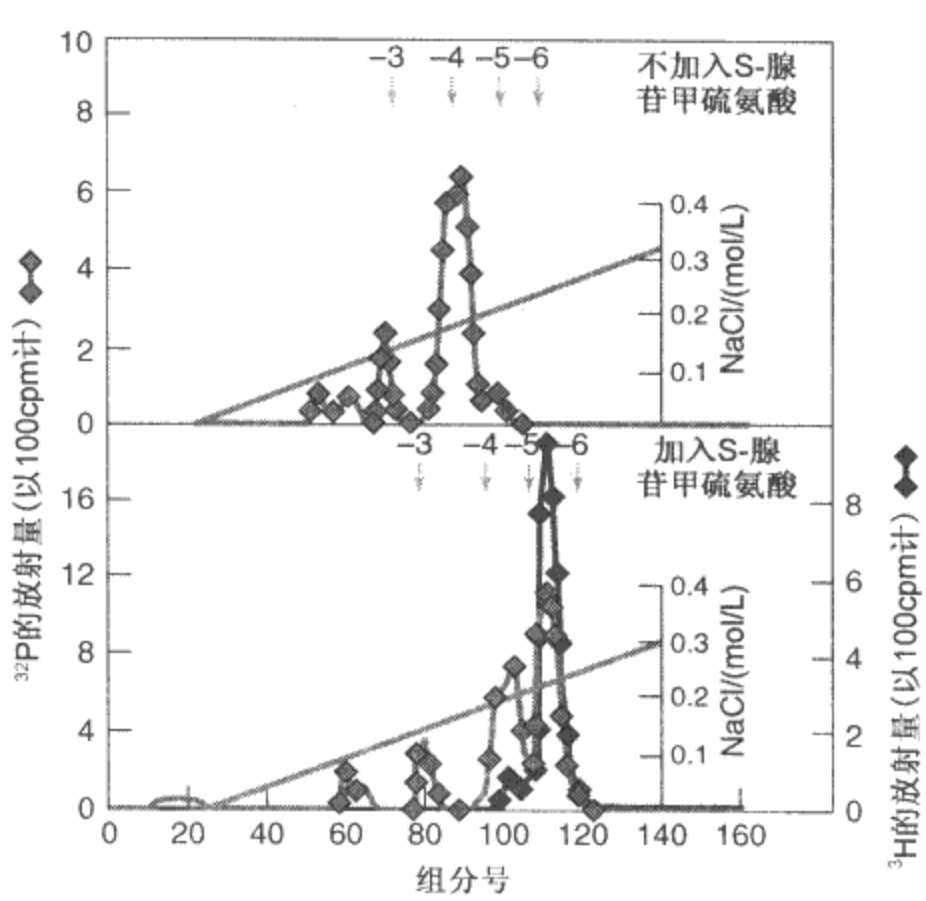


图 15.1　DEAD-纤维素层析法分离牛痘病毒 mRNA 的帽子结构。Wei 和 Moss 让牛痘病毒在具有［β，γ-^{32}P］ATP 的情况下合成帽子结构，此过程中不加入（a）或加入（b）［^{3}H］标记的 S-腺苷甲硫氨酸。之后用 KOH 消化合成的 RNA，用 DEAD-纤维素层析法分析消化产物。将不同组分的^{3}H（蓝线）和^{32}P（红线）放射量（每分钟计数）进行了绘图，对不同片段的 NaCl 浓度（绿线）也进行了绘图。图上方是不同片段的位置及所带的净电荷。（*Source*：Adapted from Wei, C. M. and B. Moss, Methylated nucleotides block 5′-terminus of vaccinia virus messenger RNA, *Proceedings of the National Academy of Sciences USA* 72（1）：318-322, January 1975.）

7-甲基化鸟苷（7-methylguanosine，**m^{7}G**）一致，表明这个帽子就是 m^{7}G。

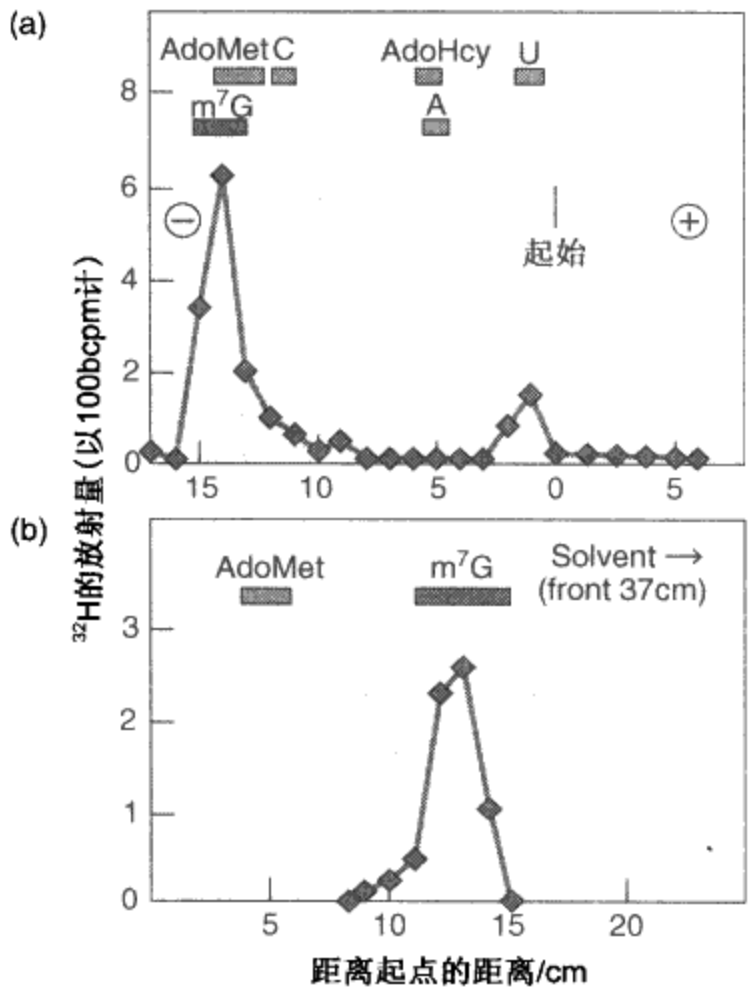


图 15.2　加帽底物（X）为 7-甲基鸟苷的证实。Miura 和 Furuichi 用磷酸单酯酶消化^{3}H 标记的加帽底物（Xp）生成 X。(a) 将消化产物与一系列分子质量标准物电泳（如 S-腺苷甲硫氨酸、m^{7}G、S-腺苷高半胱氨酸、腺苷、尿苷）。由于电泳未能分离 S-腺苷甲硫氨酸和 m^{7}G，他们又用纸层析法（b）进行分析。放射性显示的底物 X 与 m^{7}G 的迁移距离一致。（*Source*：Data from Furuichi, Y. and K.-I. Miura, A blocked structure at the 5′ terminus of mRNA from cytoplasmic polyhedrosis virus. *Nature* 253：375, 1975.）

磷酸二酯酶消化后的另外一个产物是 pAm（2′-*O*-甲基单磷酸腺苷）。因此，帽子中 m^{7}G 是与 pAm 相连的。这种连接的本质是什么？以下两点证明它们之间通过三磷酸连接：①帽子中的 GTP 只保留了 α-磷酸，没有 β-或 γ-磷酸；②帽子中的 ATP 保留了 α-磷酸和 β-磷酸。因此，一个磷酸基团来自 GTP，还有两个来自起始 RNA 合成的 ATP，在加帽的 m^{7}G 与下一个核苷之间总共有 3 个磷酸基团相互连接。此外，由于 ATP 和 GTP 在其 5′端都有磷酸基团，二者之间很可能以 5′-5′方式连接。

那么如何解释呼肠孤病毒中帽子 5′端所带的 5 个负电荷呢？图 15.3 这一问题进行了分析。其中 3 个负电荷来自 m^{7}G 与倒数第二个核苷之间的三磷酸。另一个负电荷源于倒数第二个核苷与倒数第三个核苷之间的磷酸二酯键（此键没有被碱裂解是因为其 2′位的羟基被甲基化了）。还有两个负电荷来自帽子末端的磷酸基团。这里总共是 6 个负电荷，但是 m^{7}G 自身提供了一个正电荷，所以最终电荷量为 −5。

其他的病毒和细胞内的 mRNA 都有相似的帽子，根据 2′位羟基被甲基化的程度划分为 3 种类型。第一种类型（**帽子 1**）如图 15.3 所示。第二种类型（**帽子 2**）还有一个 2′位羟基甲基化核苷酸（一排有 2 个）。第三种类型（**帽子 0**）没有 2′位羟基被甲基化的核苷酸。第二种帽子序列只出现在真核生物的细胞内，第一种则在细胞和病毒 RNA 中均有，而第三种只存在于某些病毒 RNA 中。我们在第 14 章中讲过，某些 snRNA 还有另一种帽子，其鸟苷上有 3 个位点被甲基化。本章我们还会再谈到这些帽子序列。

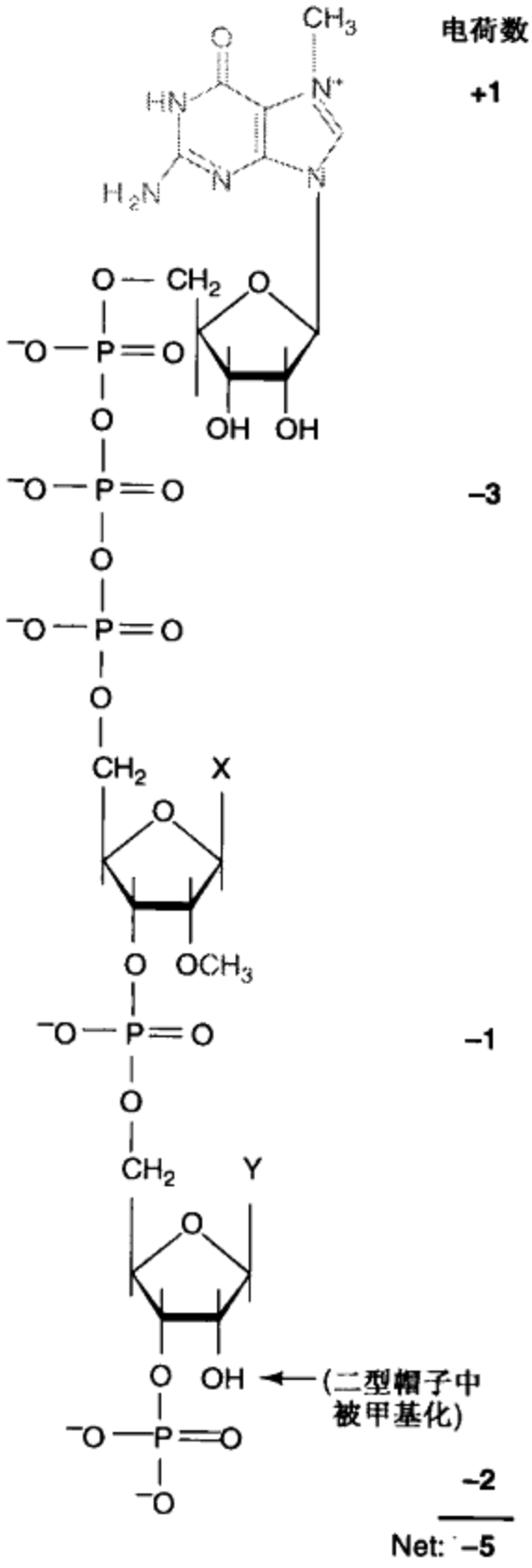
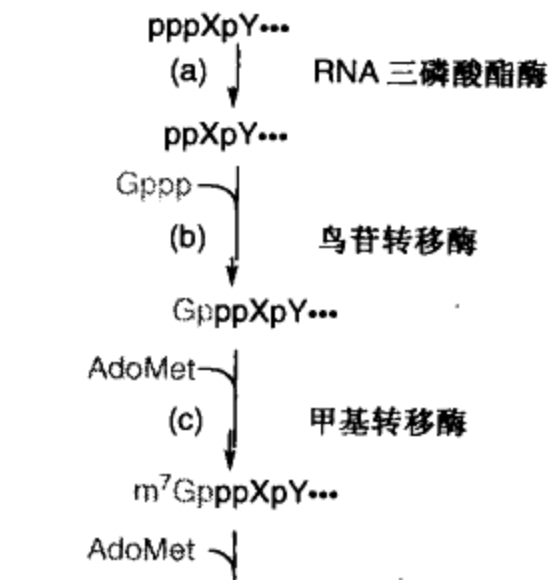


图 15.3　呼肠孤病毒帽子结构（帽子 1）所带的电荷。m^7G（带有红色甲基的蓝色鸟苷）提供一个正电荷，焦磷酸连接提供 3 个负电荷，磷酸二酯键提供一个负电荷，末端的磷酸基团提供 2 个负电荷。因此其净电荷是－5。与 Y 碱基相连的核糖上 2′-OH 在帽子 2 结构中会被甲基化。

帽子的合成

为了明确帽子是如何合成的，Moss 小组和 Furuchi、Shatkin 小组在体外进行了模式底物的加帽实验。他们分别分离了牛痘病毒和呼肠孤病毒的壳内成分（core），以此来提供加帽所需的蛋白酶。这两种人类病毒都是在宿主细胞的胞质中进行复制，不需要进入其细胞核内。因此，它们的壳内就应该含有自己的转录和加帽系统。如图 15.4 所示，在这两种病毒内所发生的事件顺序是一致的。第一步（图 15.4a），核苷磷酸水解酶（也称 **RNA 三磷酸酶**）将正在合成的 RNA 分子 5′端的 γ-磷酸水解掉，生成一个焦磷酸；第二步（图 15.4b），鸟苷酸转移酶将来自 GTP 的 GMP 连接到 RNA 末端的焦磷酸上，形成一个 5′-5′-三磷酸；第三步（图 15.4c），甲基转移酶将 S-腺苷甲硫氨酸的甲基转移到末端鸟苷酸的 7-N 位置；第四步（图 15.4d），另外一个甲基转移酶再从另一个 AdoMet 分子上将甲基转移到倒数第二个核苷的 2′-OH。



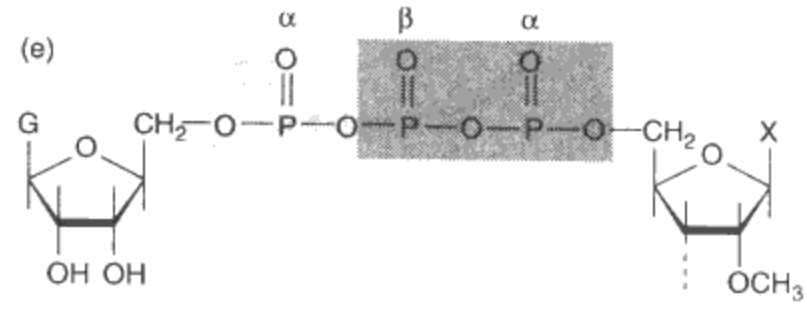


图 15.4　加帽过程中事件的顺序。（a）RNA 三磷酸酶从延伸中的 RNA 的 5′端切去 γ-磷酸基；（b）鸟苷转移酶从 GTP（蓝色）上将 GMP 转移到 RNA 上形成三磷酸基连接，封闭 RNA 的 5′端；（c）甲基转移酶从腺苷甲硫氨酸上转移甲基基团（红色）到封闭鸟苷的 7 位氮上；（d）另外一个甲基转移酶从腺苷甲硫氨酸上转移一个甲基基团（红色）到倒数第二的核苷酸的 2′-OH 上，形成帽子 1，为了形成帽子 2，在下一个核苷酸上重复步骤（d）进行甲基化；（e）在三磷酸基团连接中，磷酸基团的来源。α 位和 β 位的磷酸基团来自起始核苷酸（XTP）用绿色标出，帽子 GTP 中的 α 位磷酸用黄色标出。

为了验证上述反应途径的正确性，他们分离了此过程中所涉及的酶及中介物。例如，Furuichi 及同事用标记的 pppGpC 为模式底物进行研究。如何知道病毒分子从底物上去掉了一个磷酸基团而将之转变为 ppGpC 呢？他们用过量的副产物（PPi）来阻止鸟苷转酯反应，使 ppGpC 积累。按照图 15.5 中的方案对这个中间产物进行检测。首先，他们采用纸电泳方法，发现其中一个重要的标记产物与 ppGpC 的迁移率一样。但不巧的是，CDP 也迁移到相同位置，所以不能明确证实标记产物为 ppGpC。于是，他们用碱性磷酸酶将产物中的 ppGpC 转换为 GpC，然后再重新电泳。这一次放射性的峰值出现在 GpC 的位置上。这是令人振奋的结果，但为了从正面证实中间产物就是 ppGpC，他们将实验（a）中所得到的峰值产物进行了离子交换树脂层析，结果发现放射性峰值产物与 ppGpC 标准物是共迁移的。这样，确定了 ppGpC 是加帽过程的中间产物。实验中 ppGpC 峰值处出现少量的^{14}C标记是因为^{14}C本身的放射性强度更低一些。

帽子结构是什么时候加上的呢？在某些病毒中，如胞质多角体病毒（cytoplasmic polyhedrosis virus，CPV），S-腺苷甲硫氨酸的缺失会抑制转录，表明其转录是依赖于加帽过程的。由此推测这种病毒的加帽是个非常早期的事件，有可能在前体 mRNA 刚开始形成第一个磷酸二酯键之后就发生了。在其他病毒如牛痘病毒中，缺少 S-腺苷甲硫氨酸，其转录过程仍然正常进行，因此这种病毒的转录和加帽没有紧密联系。

与 CPV 和牛痘病毒不同，腺病毒是在核内进行复制的，因而有可能利用寄主的加帽系统进行加帽。因此，腺病毒应该可以告诉我们真核细胞前体 mRNA 的加帽过程。James Darnell 及同事证实，腺病毒的加帽过程发生在转录的早期。他们检测了腺病毒帽子和 mRNA 前体分子中前 12 个左右核苷酸中所掺入的［^{3}H］腺苷的量。首先，在腺病毒感染的后期加入［^{3}H］腺苷来标记前体 mRNA 帽子（$m^{7}GpppA$ 中的粗体 A）和其他位置的腺苷。然后用梯度离心法将小分子前体 mRNA 的大片段剔除，再以含有中后期转录物起始位点的限制性片段为探针进行杂交。能够被杂交上的任何小分子 RNA 都有可能是刚起始合成的 RNA，不仅仅只是成熟 RNA 的降解片段。从杂交分子上将新生的 mRNA 洗脱下来，看是否已被加帽。果然这些 mRNA 分子已被加上了帽子结构，而且不含有三磷酸腺苷（pppA），该分子只在没有加帽的 mRNA 末端出现。实验结果表明，腺病毒 mRNA 在其转录的中后期或延伸到 70 个核苷酸之前就已经加帽了。目前所公认的是，在真核细胞内，加帽发生得更早，可能在前体 mRNA 延伸到 30 个核苷酸之前。

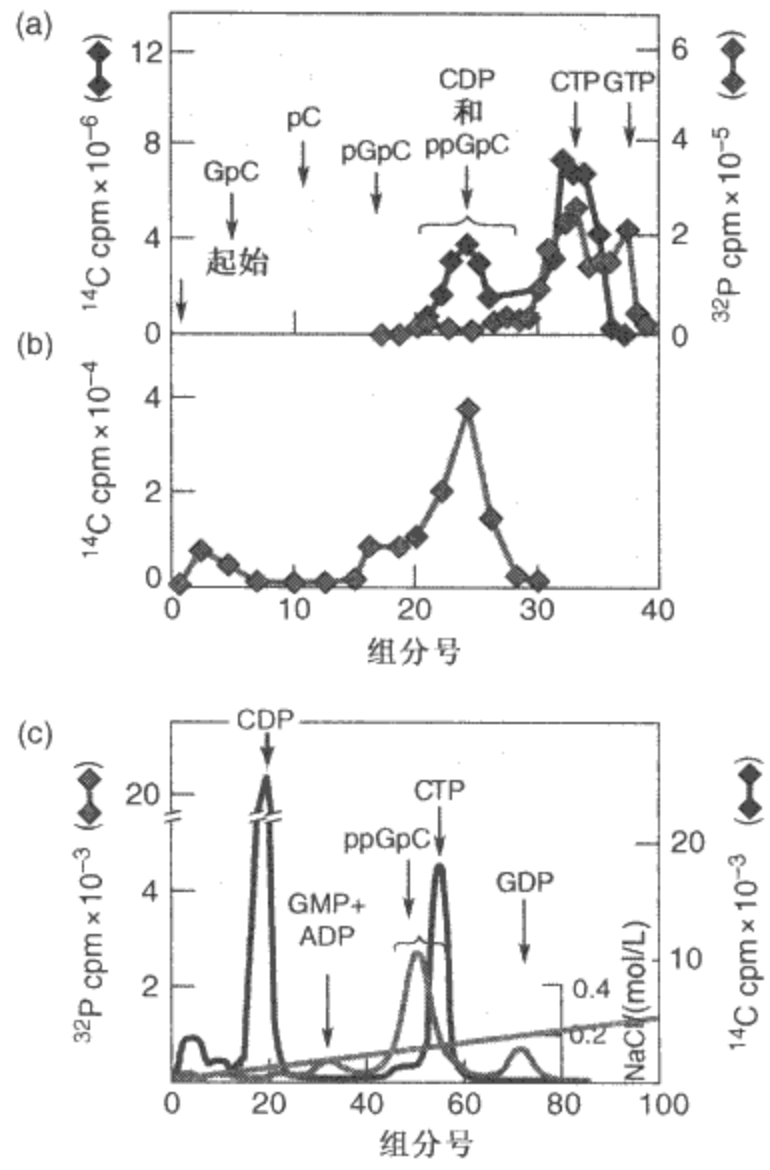


图 15.5　验证呼肠孤病毒帽子合成的中间产物为 ppGpC。（a）第一次纯化步骤。Furuichi 及同事在呼肠孤病毒中加入［^{14}C］CTP 和［^{32}P］GTP 来标记帽子以及加帽过程的中间产物。之后，将标记物跟图上方所列的已知化合物一起进行纸电泳。一种放射性的中间产物（弧形曲线部分）与 ppGpC 和 CDP 迁移率相同。（b）ppGpC转换为 GpC。将（a）中弧形曲线所对应的物质用碱性磷酸酶消化，使 ppGpC 转换为 GpC，然后再进行电泳。这次，一个主峰与 GpC 共迁移。（c）ppGpC 的阳性验证。将（a）中弧形曲线所对应的物质用离子交换树脂层析，所用已知化合物标于图顶部。^{32}P 标记的主峰与 ppGpC 是吻合的。（*Source*：Adapted from Furuichi Y., S. Muthukrishnan, J. Tomasz, and A. J. Shatkin, Mechanism of formatioin of reovirus mRNA 5′-terminal blocked and methylated sequence $m^{7}GpppG^{m}pC$. *Journal of Biological Chemistry* 251：5051，1976.）

小结　帽子是分步合成的。首先，RNA 焦磷酸酶将前体 mRNA 末端的磷酸基团去掉，之

后鸟苷转移酶在此末端加上 GMP（来源于 GTP）。然后，两个甲基转移酶分别将鸟苷第 7 位的 N 和倒数第二个碱基 2′-OH 甲基化。这一系列事件发生在 RNA 链延伸到 30 个核苷酸之前的转录早期。

帽子的功能

帽子结构至少有 4 种功能：①保护 mRNA 免遭降解；②加强 mRNA 的可译性；③增强 mRNA 从细胞核到细胞质的转运；④提高 mRNA 的剪接效率。在本节我们先讨论帽子结构的前 3 个功能，下一章再讨论其第 4 个功能。

保护功能 帽子结构是通过三磷酸键与随后的 mRNA 连接的，这种磷酸键在 RNA 的其他部位没有发现过。因此，帽子结构可以通过防止从 5′端开始的 RNase 对 mRNA 分子的外切作用来行使保护功能，这种 RNase 对焦磷酸键无能为力。事实上，已有明确的证据表明，帽子可以防止 mRNA 分子的降解。

Furuichi、Shatkin 及同事于 1977 年发表文章，证明加帽的呼肠孤病毒 RNA 比未加帽的要稳定得多。他们分别对新合成的用 m^7GpppG 加帽的、用 GpppG 加帽的以及未加帽的 RNA 进行标记，然后将这 3 种 RNA 分别注入非洲爪蟾的卵母细胞。8h 后用甘油密度梯度超速离心进行分离和分析。结果发现，呼肠孤病毒 RNA 以 3 种大小存在，分别为大（l）、中（m）、小（s）3 种片段。图 15.6a 显示甘油密度梯度超速离心对这 3 种大小 RNA 的分离结果。对含有不同 5′端的 3 种 RNA 都进行了相同的分析，并没有发现明显的不同。实验中 3 条大小不同的条带总是清晰可见。图 15.6b 所显示的是非洲爪蟾卵母细胞放置 8h 之后的结果，3 种不同 5′端的 RNA 都有降解，但这种降解在未加帽的 RNA 中更为明显。由此看来，非洲爪蟾卵母细胞虽然含有可降解病毒 RNA 的核酸酶，但帽子可以提供一定的保护功能。

可译性 帽子的另一个重要功能是保证 mRNA 的可译性。在第 17 章我们将看到，真核细胞的 mRNA 进入核糖体进行翻译是通过帽子结合蛋白对帽子的识别而实现的。如果没有帽子结构，帽子结合蛋白就不能结合上去，mRNA 的翻译效率就会非常低。Daniel Gallie 在一次活体实验中，描述了帽子对翻译的促进作用。Gallie 分别将有帽和无帽，同时有和无 Poly（A）的萤火虫萤光素酶 mRNA 导入烟草细胞。由于萤光素酶在有荧光素和 ATP 的情况下能够发光，所以很容易检测。表 15.1 的结果表明，3′端的 poly（A）和 5′端的帽子都能协同增强萤光素

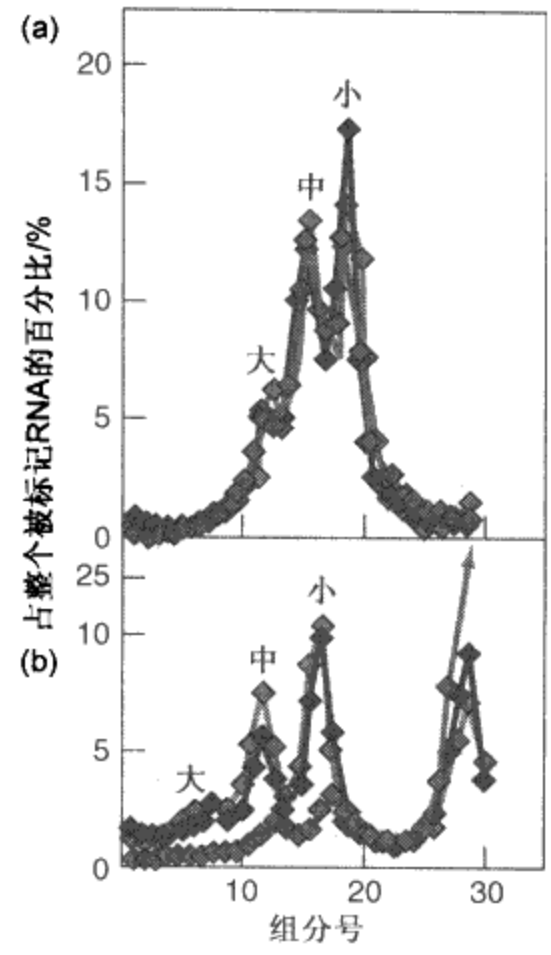


图 15.6 帽子结构对呼肠孤病毒 RNA 稳定性的影响。（a）新合成的 RNA。对已经加帽的（绿色）、封堵的（蓝色）未加帽的（红色）呼肠孤病毒 RNA 进行标记，并对其进行甘油密度梯度超离心分析。3 种不同大小的 RNA 片段分别被标为大（l）、中（m）、小（s）。（b）RNA 分子在非洲爪蟾卵母细胞内培养的结果。将以上含有不同 5′端的 RNA 注射到非洲爪蟾卵母细胞中，8h 后分离，并用（a）中的离心方法进行分析。颜色标注与（a）相同。（*Source*：Adapted from Furuichi，Y.，A. LaFiandra，and A. J. Shatkin，5′-terminal struture and mRNA stability. *Nature* 266：236，1977.）

酶的稳定性和翻译效率。对于已加帽的 mRNA，poly（A）尾可将其翻译效率提高 21 倍，但加帽过程却可将含有 poly（A）尾的 mRNA 的翻译效率提高 297 倍。当然，mRNA 自身的稳定性对结果的影响已考虑在内，但其影响并不是很大。

RNA 的运输 帽子结构同时还有利于成熟 RNA 向核外的运输。Jörg Hamm 和 Iain Mattaj 在分析了 U1 snRNA 的转运之后得出这一结论。大多数 snRNA，包括 U1 snRNA 在内，都是由 RNA 聚合酶 II 转录的，其转录物在细胞核内被加上单甲基化的 m^7G 帽子。这些转录物被转运到细胞质，在这里与一些蛋白质结合形成 snRNP 复合物，同时它们的帽子被进一步甲基化为三甲基鸟苷（$m^{2,2,7}$G）。然后再次进入到核内完成剪切及其他加工过程。U6 snRNA 是一个例外，它由聚合酶 III 合成，且没有

帽子结构，保留了自己的末端焦磷酸，并留在核内。Hamm 和 Mattaj 想弄清楚如果 U1 snRNA 被聚合酶 III 而不是聚合酶 II 合成，情况会如何？如果它不被加帽，而且留在核内，就表明帽子结构对于 RNA 向核外转运是重要的。

表 15.1　poly（A）和帽子结构对烟草原生质体萤光素酶 mRNA 翻译过程的协同作用

mRNA	萤光素酶 mRNA 半衰期/min	萤光素酶活性/（光单位/mg 蛋白）	poly（A）尾对酶活性的相对影响	帽子对活性的相对影响
未加帽的				
没有 poly（A）	31	2941	1	1
含有 poly（A）	44	4480	1.5	1
加帽的				
没有 poly（A）	53	62 595	1	21
含有 poly（A）	100	1 331 917	21	297

Source: Gallie, D. R., The cap and poly（A）tail function synergistically to regulate mRNA translational efficiency, *Genes & Development* 5: 2108-2116, 1991. Copyright © Cold Spring Harbor, NY. Reprinted by permission.

于是，Hamm 和 Mattaj 将非洲爪蟾 U1 snRNA 基因放在人的 U6 snRNA 启动子之后，使它由聚合酶 III 转录。之后，将构建好的 U1 snRNA 注入爪蟾卵母细胞核中，同时还注入一种标记的核苷酸和爪蟾的 5S rRNA 基因作为内参。实验中还加入了 1μg/ml 的 α-鹅膏蕈碱抑制 RNA 聚合酶 II 的活性，以此确保 U1 基因的转录物不会由聚合酶 II 合成。除了野生型 U1 基因外，还用几种突变型 U1 基因进行了分析，这些突变发生在编码蛋白质结合位点的区域。在失去与胞质特定蛋白质结合的能力后，突变转录物在转运到胞质后不能再回到细胞核。注射 12h 后，Hamm 和 Mattaj 剥离了卵母细胞的细胞质和细胞核，并对两种组分的标记产物进行了电泳检测。比较了由 RNA 聚合酶 II 合成并加帽了的 U1 snRNA 和由 RNA 聚合酶 III 合成但未加帽的 U1 snRNA 在细胞中的定位。

结果，所有聚合酶 III 合成的未加帽 U1 snRNA 都被保留在核内。而聚合酶 II 合成的 U1 snRNA 被转运到胞质。这些结果证实了加帽对于 U1 snRNA 往核外运输是必需的这一假设。

稍后我们在这一章将看到，帽子结构对于前体 mRNA 的正确剪接也是必不可少的。

小结　帽子结构有 4 个功能：①保护 mRNA 不被降解；②增加 mRNA 的可译性；③将 mRNA 转运到核外；④使前体 mRNA 正确剪接。

15.2　多聚腺苷酸化

我们已经知道 hnRNA 是 mRNA 的前体。二者存在如下联系：它们的 3′端共有一个独特结构，即被称为 poly（A）的 AMP 残基长链。rRNA 和 tRNA 都不具有 poly（A）尾。给 RNA 添加 poly（A）的过程称为**多聚腺苷酸化**（polyadenylation）。我们先了解一下 poly（A）的性质，然后学习多聚腺苷酸化的过程。

poly（A）

James Darnell 小组做了大量关于 poly（A）和多聚腺苷酸化的早期研究。为了从 HeLa 细胞的 mRNA 中得到 poly（A）片段，他们用 RNase A 和 RNase T1 两种酶切割 mRNA。其中，RNase A 在 C 和 U 后酶切 RNA，RNase T1 在 G 后酶切 RNA，即剪接除了 A 之外的所有核苷酸，只留下纯的腺嘌呤片段。之后电泳细胞质和细胞核的 poly（A），确定其大小。图 15.7 显示，两种 poly（A）片段都在比 5S rRNA 慢一些的位置具有主峰，大概是 7S。他们估计这些片段有 150～200 个碱基。在该实验中观察到的 poly（A）片段只被标记了 12min，所以它们是新合成的。值得注意的是，新合成的核内和细胞质 poly（A）片段的大小没有多大差别。但在本章后面我们将看到细胞质 poly（A）会逐渐变短。分析不同生物的 poly（A）尾发现，新合成 poly（A）尾大小平均约为 250nt。

poly（A）是从 mRNA 或 hnRNA 的 3′端进行延伸的。在一个从 3′端降解的 RNase 作用下，poly（A）可以从 mRNA 或 hnRNA 的 3′端很快释放出来。因此，用 RNase 完全降解 poly（A）所得产物是 1 分子的腺苷和约 200 个分子的 AMP。图 15.8 表明得到这一结果要求 poly（A）位于 mRNA 分子

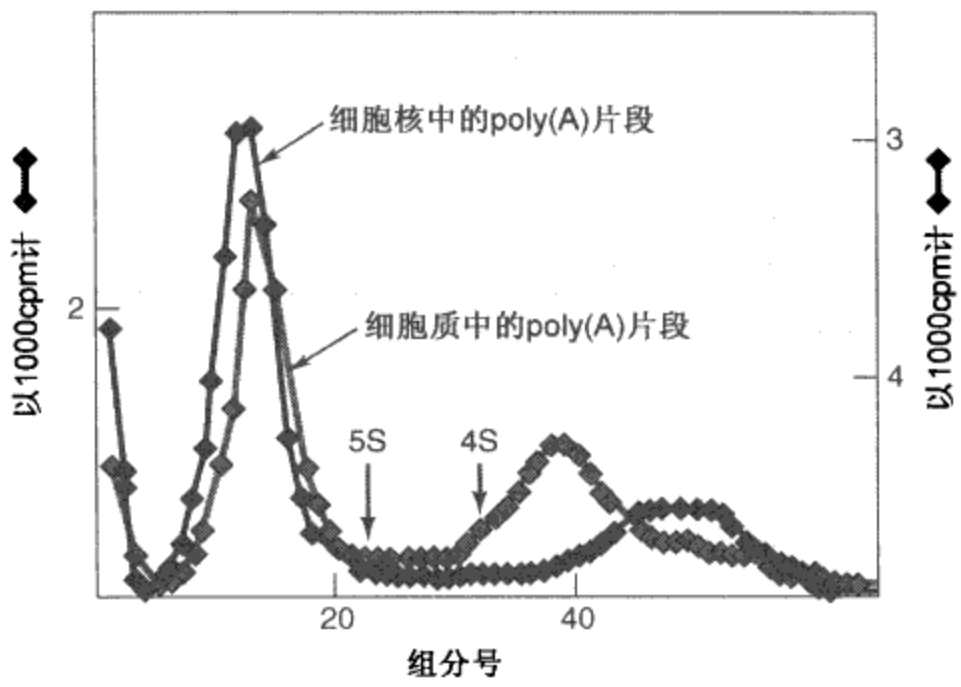


图 15.7　poly（A）的大小。Sheiness 和 Darnell 从 HeLa 细胞的细胞核和细胞质中分别分离放射性标记的 hnRNA（蓝色）和 mRNA（红色）。然后用 RNase A 和 RNase T1 处理这些 RNA 以释放 poly（A），收集碎片，用闪烁计数法测量放射强度（第 5 章）。用 4S tRNA 和 5S rRNA 作为分子大小标记。相应的分子长度为 200nt。（*Source*：Adapted from Sheiness, D. and J. E. Darnell, Polyadenylic acid segment in mRNA becomes shorter with age. *Nature New Biology* 241：267, 1973.）

的 3′端。该实验也进一步证明 poly（A）的长度约为 200nt。

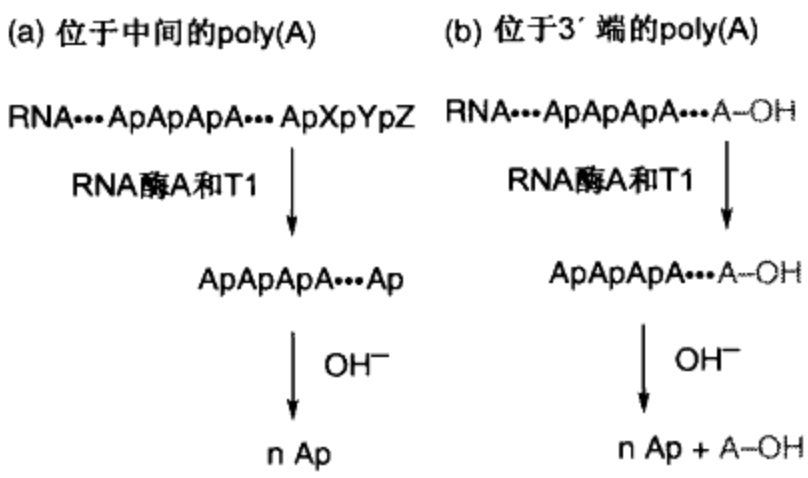


图 15.8　在 hnRNA 和 mRNA 3′端发现 poly（A）。（a）内部 poly（A）。如果 poly（A）存在于 RNA 分子内部，RNase A 和 RNase T1 消化时产生的 poly（A），在其 3′端将带有一个磷酸基团。碱基水解只能产生 AMP。（b）poly（A）存在于 hnRNA 和 mRNA 的 3′端。因为 poly（A）存在于这些 RNA 分子的 3′端，RNase A 和 RNase T1 消化时在 3′端产生没有磷酸化的腺嘌呤。碱基水解产生 AMP，另外有 1 分子腺苷。AMP 与腺苷的比率为 200，提示 poly（A）的长度是 200nt。

我们知道 poly（A）不是由 DNA 转录而来，因为基因组中没有足够长的一串 T 来编码它。举例来讲，在数千个已经测序的真核生物基因的尾部没有发现 T 串序列。而且，放线菌素 D 虽然可抑制 DNA 转录过程，但是并不能抑制多聚腺苷酸的合成。所以，poly（A）一定是在转录后加上去的。事实上，细胞核内有一种酶叫做**腺嘌呤聚合酶**［poly（A）polymerase，PAP］，它可以在 mRNA 前体上每次添加一个 AMP 残基。

由于在 hnRNA 上发现有 poly（A），所以我们知道 poly（A）尾是加在 mRNA 前体上的，甚至是一些特殊的尚未剪接的 mRNA 前体上（如 15S 鼠球蛋白 mRNA 前体）也含有 poly（A）。但是在本章后面，我们将看到前体 mRNA 上一些内含子的剪切要早于多聚腺苷酸化。一旦一个 mRNA 进入细胞质，其 poly（A）就开始降解，换句话说，poly（A）不可避免地受到 RNase 的降解，然后又被细胞质中的 poly（A）聚合酶重新合成。

小结　绝大多数真核生物的 mRNA 及其前体在 3′端都有一段约 250nt 的多聚腺苷酸残基链。这段 poly（A）是在转录之后被 poly（A）聚合酶加上的。

poly（A）的功能

绝大多数 mRNA 都有 poly（A），但组蛋白 mRNA 是个特例，它在没有 poly（A）尾的情形下仍能行使自己的功能。虽然有这个例外，但真核生物几乎普遍存在的 poly（A）到底发挥什么作用呢？研究表明，poly（A）具有保护 mRNA 免受降解的功能，poly（A）对它所连接的 mRNA 的翻译有促进作用，还有证据表明 poly（A）在 mRNA 的剪接及转运出核时发挥一定的作用。在此，我们先寻求 poly（A）在 mRNA 稳定性和可译性方面的证据，在本章的最后我们再讨论剪接和转运功能。

保护 mRNA　为了检验 poly（A）对 mRNA 的稳定作用，Michel Revel 及同事将有 poly（A）尾和无 poly（A）尾的球蛋白 mRNA 注入爪蟾卵母细胞中，之后的 2 天中在不同时间测定球蛋白的合成速率。起初他们发现没有什么差异，但 6h 之后，无 poly（A）尾的 mRNA［poly（A）$^-$ RNA］不能再继续翻译下去，而有 poly（A）尾的 mRNA［poly（A）$^+$ RNA］依旧可以有效地翻译（图 15.9）。对这

一现象最简单的解释就是，poly（A）$^+$ RNA 比 poly（A）$^-$ RNA 有较长的稳定时间，因此 poly（A）是个保护剂。但我们也将看到，有些 mRNA 的 poly（A）尾并不具有保护作用。但可以明确的是，poly（A）尾在促进 mRNA 的翻译效率方面具有更为重要的作用。

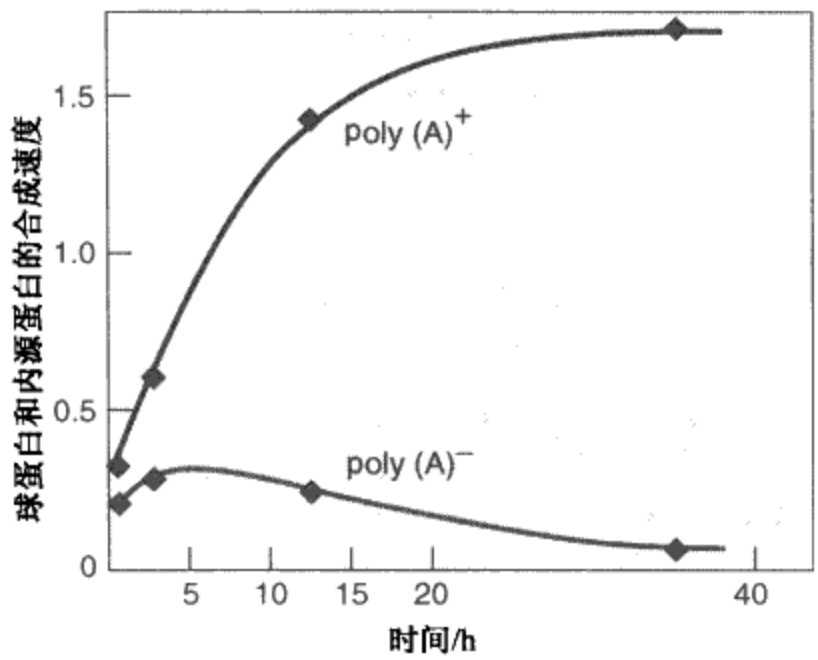


图 15.9　poly（A）$^+$（蓝色）与 poly（A）$^-$（红色）球蛋白 mRNA 翻译的时间进程。Revel 及同事绘制的比率图，以整合到球蛋白和内源蛋白中的放射性对标记时间的中点作图。（*Source*：Adapted from Huez, G., G. Marbaix, E. Hubert, M. Leclereq, U. Nudel, H. Soreq, R. Solomon, B. Lebleu, M. Revel, and U. Z. Littauer, Role of the polyadenylate segment in the translation of globin messenger RNA in *Xenopus* oocytes. *Proceedings of the National Academy of Sciences USA* 71（8）：3143-3146, August 1974.）

增强 mRNA 的可译性　有多个证据表明，poly（A）尾可增强 mRNA 的翻译能力。真核生物 mRNA 翻译中有一个结合蛋白是 **poly（A）结合蛋白 I**［poly（A）-binding protein I，PAB I］。mRNA 翻译时与该蛋白质结合可增强翻译效率。支持该假说的实验证据是：在体外翻译体系中加入过量 poly（A）可抑制加帽的、多聚腺苷酸化的 mRNA 的翻译。这个实验暗示过量 poly（A）与 mRNA 上的 poly（A）竞争某个重要因子，推测是 PAB I。无 PAB I 时，mRNA 不能很好地翻译。那么我们可以进一步得出这样的结论，即 poly（A）$^-$ RNA 如果不能结合 PAB I，就不能有效地进行翻译。

为了检验这个假说，David Munroe 和 Allan Jacobson 比较了两个人工合成的有 poly（A）和无 poly（A）的 mRNA 在兔网织红细胞提取物中的翻译速率。他们克隆了兔 β-球蛋白基因和疱疹性口炎病毒的 N 基因并连接到质粒中，在噬菌体 SP6 启动子调控下，用 SP6 RNA 聚合酶进行体外转录，分别获得了这两个基因的 mRNA［兔 β-球蛋白（RBG）mRNA 和疱疹性口炎病毒的 N 基因（VSV. N）mRNA］。在克隆后和转录前，用末端转移酶和 dTTP 以不同反应时间，分别给每个基因添加 poly（T）尾，使这些人工合成的 mRNA 具有不同长度的 poly（A）尾。

他们在网织红细胞提取物中分别测定了 poly（A）$^+$ 和 poly（A）$^-$ mRNA 的稳定性和可译性。图 15.10 显示了帽子和 poly（A）尾对 VSV. N mRNA 稳定性和可译性的影响。对于有帽子或无帽子 mRNA 而言，在有 poly（A）尾时要比无 poly（A）尾时翻译得好。进一步实验表明，多聚腺苷酸化对这两种 mRNA 的稳定作用并没有差异。他们对这一结果的解释是：poly（A）对 mRNA 翻译的增强作用不是由于加强了 mRNA 的稳定性，而是由于加强了翻译过程本身。如果是这样，那么，poly（A）到底加强了翻译过程的哪一步呢？以上研究表明，这一作用发生在翻译的起始阶段，即 mRNA 和核糖体结合时。在第 17 章我们将看到，许多核糖体有序地结合到真核细胞 mRNA 的头部并依次读取 mRNA 的信息。一条 mRNA 链一次结合数个核糖体进行翻译的结构称为多聚核糖体（polysome）。Munroe 和 Jacobson 认为 poly（A）$^+$ mRNA 比 poly（A）$^-$ mRNA 更容易成功地形成多聚核糖体。

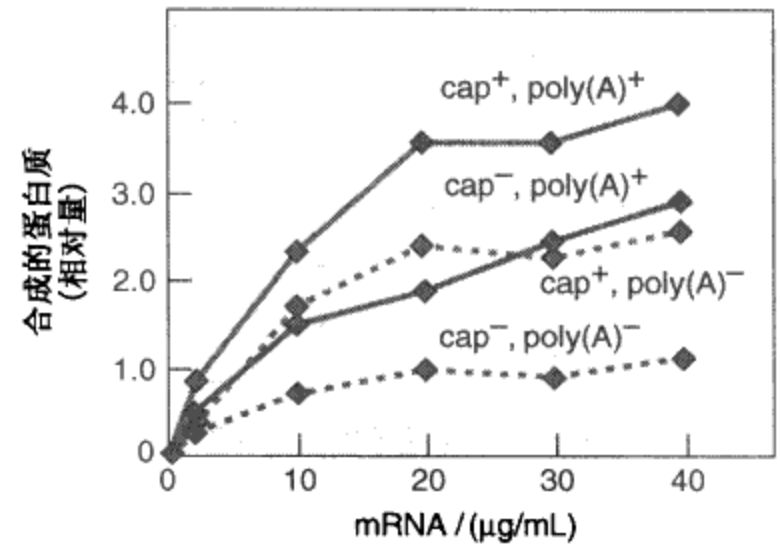


图 15.10　多聚腺苷酸化对 mRNA 可译性的影响。Munroe 和 Jacobson 将 VSV. N 的 mRNA 和 ^{35}S 标记的甲硫氨酸在兔网织红细胞提取物中孵育。RNA 分子有加帽的（绿色）和未加帽的（红色），有 poly（A）$^+$（实线）或者无 poly（A）$^-$（虚线）。在蛋白质合成 30min 之后，电泳标记产物，通过定量荧光检测法测量新合成蛋白的放射活性。poly（A）对加帽和未加帽 mRNA 的可译性都有促进作用。（*Source*：Adapted from Munroe, D. and A. Jacobson, mRNA poly（A）tail, a 3′ enhancer of a translational initiation. *Molecular and Cellular Biology* 10：3445, 1990.）

Munroe 和 Jacobson 测定了标记过的 mRNA 掺入多聚核糖体的过程。他们用^{32}P 标记含 poly（A）的 mRNA，用^{3}H 标记没有 poly（A）的 mRNA，接着将这些 mRNA 在网状细胞提取液中孵育，然后用蔗糖梯度离心法提取多聚核糖体。图 15.11a 显示带有 poly（A）的 VSV. N mRNA 比不带 poly（A）的 VSV. N mRNA 更容易与核糖体结合。在平行实验中，RBG mRNA 也显示了相同的结果。图 15.11b 显示连接到 RBG mRNA 的 poly（A）的长度对多聚核糖体形成程度的影响。可以发现，随着 poly（A）从 5nt 到 30nt 的不断加长，相应的核糖体的形成也在不断地加强。

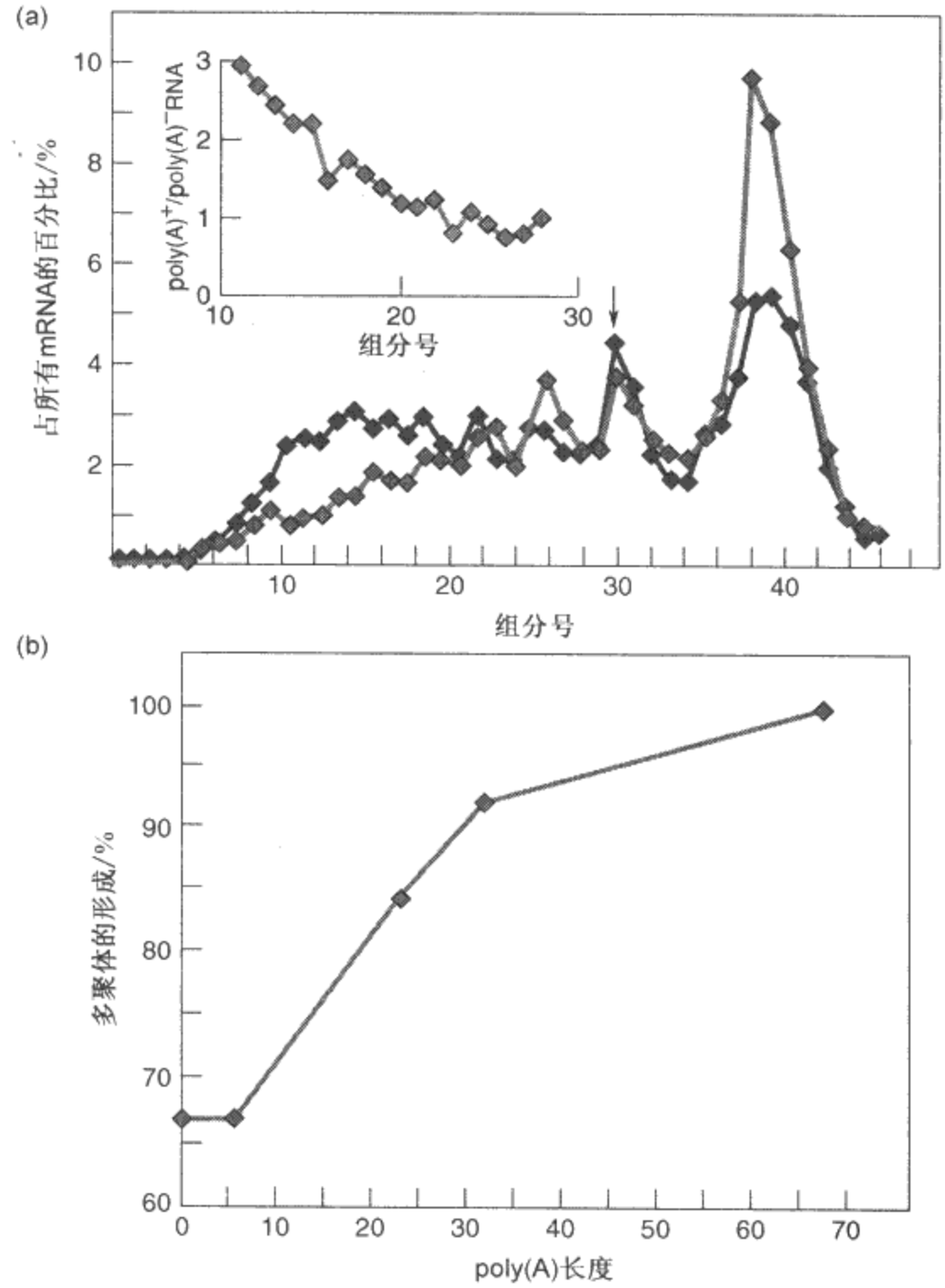


图 15.11　多聚腺苷酸化对募集 mRNA 形成多聚核糖体的作用。（a）多聚核糖体分布图。Munroe 和 Jacobson 分别将^{32}P 标记的 poly（A）$^{+}$（蓝色）mRNA 和^{3}H 标记的 poly（A）$^{-}$（红色）mRNA 与兔网织红细胞提取物混合，然后用蔗糖梯度离心分离多聚核糖体与单体。箭头指示单体的峰值，其左边是多聚体片段，可以看到二聚体、三聚体甚至多聚体的峰值。有 poly（A）尾的 mRNA 显然与多聚核糖体结合得更好，尤其是与高聚体。插图显示组分 11～28 中 poly（A）$^{+}$与 poly（A）$^{-}$ mRNA 的比值。结果再次证明 poly（A）$^{+}$mRNA 与核糖体结合的优先性。(b) 不同长度 poly（A）对 VSV. N mRNA 和核糖体形成多聚核糖体的效率影响，长度为 68nt 的 poly（A）尾形成多聚核糖体的效率达 100%。（*Source*：Adapted from Munroe, D. and A. Jacobson, mRNA poly（A）tail, a 3′enhancer of a translational initiation. *Molecular and Cellular Biology* 10：3447-8，1990.）

Munroe 和 Jacobson 关于 poly（A）并不影响 mRNA 稳定性的发现似乎与 Revel 小组早期的研究结果有矛盾。究其原因可能是 Revel 的研究是通过完整的爪蟾卵细胞完成的，而 Munroe 的研究是用无细胞体系完成的。在本章前面的表 15.1 显示，poly（A）激发了萤光素酶 mRNA 的转录，poly（A）对该 mRNA 稳定性的影响最多是加倍，而萤光素酶的产量增加了 20 倍。因此，这一系统也显示 poly（A）似乎在增强翻译方面起更重要的作用。

在第 17 章我们将看到 poly（A）如何在保护 mRNA 的同时促进其翻译的。简而言之，poly（A）可以结合胞质中的 poly（A）结合蛋白，这些蛋白质同时又与翻译起始因子（eIF4G）结合，而 eIF4G 因子又可以结合帽子上的帽结合蛋白。这样就将 3′端的 poly（A）和 5′端的帽子拉到一起，有效地将 mRNA 环化起来。这种两端都结合有蛋白质分子的环状 mRNA 比裸露的线状的 mRNA 要稳定得多。同时环状的 mRNA 也更容易进入翻译步骤，部分原因是将 mRNA 环系在一起的 eIF4G 可以帮助将核糖体召集到 mRNA 上。

小结 poly（A）可以同时加强 mRNA 的稳定性和翻译能力，这两个方面的相对重要性因不同的系统而不同。至少在兔网状细胞提取液中，poly（A）通过辅助 mRNA 募集到多聚核糖体而加强 mRNA 的翻译作用。

多聚腺苷酸化的基本机制

从逻辑上我们很容易认为，poly（A）聚合酶要等到一个转录物完全合成之后再将 poly（A）加到 RNA 的 3′端，但事实并非如此。相反，多聚腺苷酸化通常涉及 mRNA 前体的剪接，甚至在转录终止之前，就把 poly（A）加到新暴露的 3′端（图 15.12）。因此，与预期相反，在 RNA 聚合酶仍在延伸 RNA 链时，加尾装置就已经定位到上游某个位置的多聚腺苷酸化信号上，切开延长中的 RNA 链，将其多聚腺苷酸化。

Joseph Nevins 和 James Darnell 最先为这一多聚腺苷酸化模式提供了证据。他们以腺病毒主要后期**转录单位**（transcription unit）为研究对象，它有几个不同的相互重叠的 mRNA，每个 mRNA 可在 5 个不同的位点进行多聚腺苷酸化。图 15.13a 是该转录单位的示意图，显示如何利用 5 个不同的多聚腺苷酸化位点产生 5 个不同的 mRNA（此图只是示意，没有按缩放比例来画）。回顾第 14 章我们知道，这些成熟 mRNA 中每个都是由 3 个相同的前导外显子剪接到不同编码区形成的。每个转录物的 poly（A）连接到编码区的 3′端。图 15.13b 是该系统中两个不

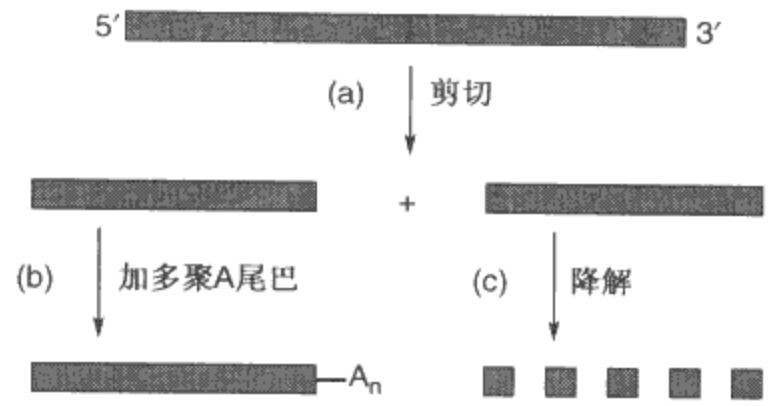


图 15.12　多聚腺苷酸化过程简图。（a）剪切。第一步是切开可能仍在转录过程中的转录物，这种剪切常发生在 RNA 的末端区（绿色）。（b）多聚腺苷酸化。poly（A）聚合酶将 poly（A）加到 mRNA 的 3′端。（c）多余 RNA 的降解，所有位于多聚腺苷酸化位点外的 RNA（红色）都被降解。

同的关于转录终止和多聚腺苷酸化之间关系的假设：①转录在多聚腺苷酸化位点的下游立即终止，接着发生多聚腺苷酸化。例如，如果基因 A 正在被表达，转录将延续到基因 A 编码区的最末端，然后转录终止，多聚腺苷酸化就在该 3′端进行；②转录至少要进行到最后的外显子末端，而多聚腺苷酸化可以在任意的多聚腺苷酸化位点发生，甚至在整个基因转录终止之前。

第一种假说认为并不总是待转录进行到最后，这一观点很容易排除。Nevins 和 Darnell 将标记的从感染晚期细胞中合成的 RNA 与主要晚期区不同区段的 DNA 片段进行杂交，如果第一个基因的初级转录物在第一个多聚腺苷酸化位点后停止转录，只有最后一个基因可以一直将转录进行到末端，那么大多数 RNA 应该与主要晚期区靠近 5′端的 DNA 片段杂交，只有少量与靠近 3′端的 DNA 片段杂交。但事实上 RNA 能均等地与所有 DNA 片段杂交。它与靠近 3′端和靠近 5′端的片段都能很好地结合。因此，主要晚期区的转录一旦开始，它将一直持续转录到该区的末端才终止。换言之，主要晚期区只含有一个终止子，且位于该区的最后面。为此，整个晚期区可以被称为一个**转录单位**（transcription unit），强调它虽含有多个基因，但作为一个整体进行转录。Nevins 和 Darnell 继续他们的研究以证实剪切和多聚腺苷酸化通常发生在转录终止之前。

在转录物被剪切和发生多聚腺苷酸化之前，转录将一直进行且远远超出多聚腺苷酸化位点这一现象看起来很浪费，因为所有延伸到多聚腺苷酸化位点之后的 RNA 都会被降解掉，没有丝毫用处。所以问题就自然而然地产生了：这种多聚腺苷酸化方式只是病毒中独有的吗？在普通细胞转录中是否也同

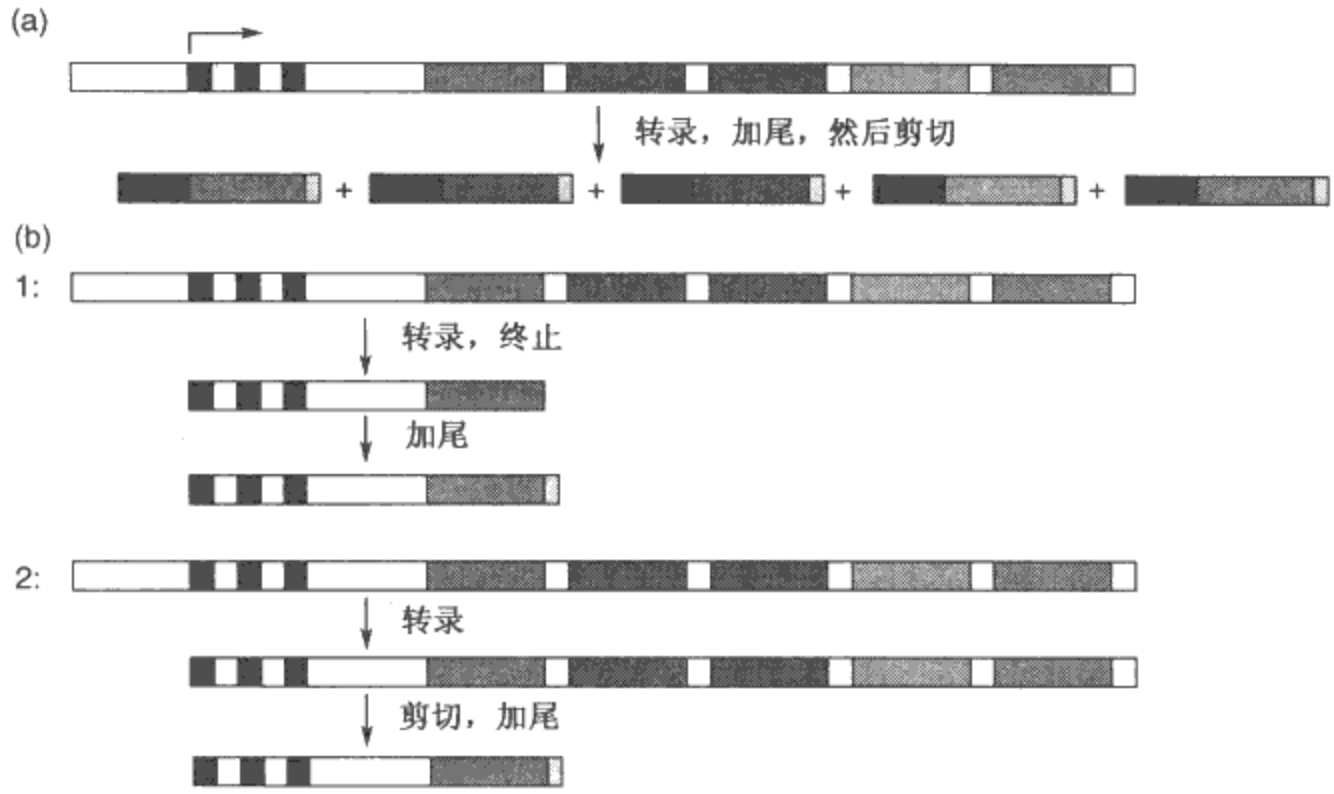


图 15.13　腺病毒主要晚期转录物的合成和多聚腺苷酸化的两个假说。（a）腺病毒主要晚期转录单元示意图。黑色前导外显子（包括帽子），编码区用不同颜色表示。转录、加帽、多聚腺苷酸化和剪接完成后，得到多重 mRNA，其编码区末端都有一个 poly（A）（黄色）。（b）腺病毒主要晚期 mRNA 多聚腺苷酸化的两个假说。每个假说只考虑第一个基因的多聚腺苷酸化。①转录在每个编码区的末端终止，多聚腺苷酸化在这些位点发生；②转录继续进行到主要晚期区的末端，初级转录物被剪接、多聚腺苷酸化在每个剪切位点发生。

样发生？为了回答这一问题，Erhard Hofer 和 James Darnell 从 Friend 鼠的红白血病（erythroleukemia）细胞中分离出带有标记的 RNA，这种细胞已被二甲基亚砜（DMSO）诱导以得到大量的球蛋白，因此其球蛋白基因会被高速转录。他们用标记的转录物与克隆的鼠 β-球蛋白基因及其下游区域的所有 DNA 片段杂交（图 15.14）。观察到对多聚腺苷酸化位点下游 500bp 区域的各种片段的杂交信号与球蛋白基因内部各种片段的杂交信号一样强。这说明转录至少延伸到多聚腺苷酸化位点下游 500bp 的位置。进一步研究发现转录最后终止在下游甚至更远的区域。因此，在细胞和病毒中，转录显然都远远地延伸到了多聚腺苷酸化位点之后。

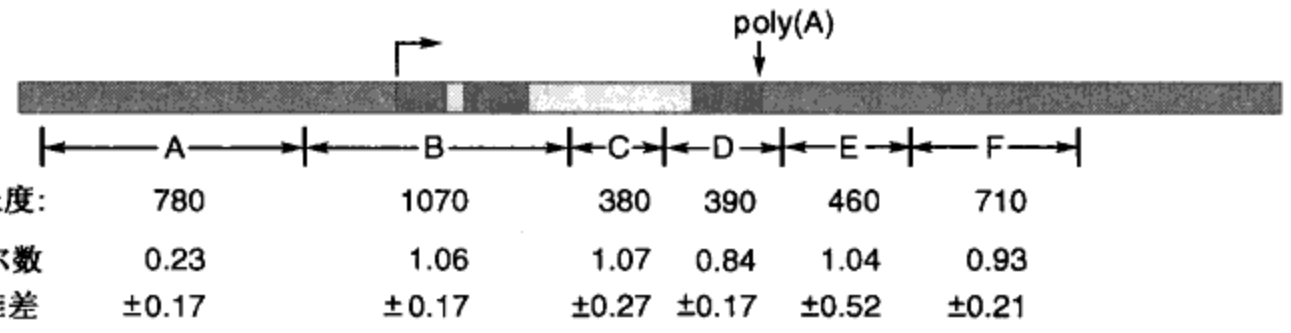


图 15.14　多聚腺苷酸化位点后的转录。Hofer 和 Darnell 从 DMSO-激发的 Friend 鼠红白血细胞中提取核酸，并把它们与 ^{32}P-UTP 共孵育以标记连续的 RNA，主要是球蛋白的 pre-mRNA。然后把这些标记的 RNA 与 DNA 片段 A～F 杂交，这些片段的大小和位置在图上部列出。RNA 与每个片段的杂交摩尔浓度以及对应的标准差（SD）在每个片段的下面标出。在物理图谱中内含子以红色显示，外显子以黄色显示。（*Source*：Adapted from E. Hofer and J. E. Darnell，The primary transcription unit of the mouse β-major globin gene. *Cell* 23：586，1981.）

小结　真核细胞基因的转录超出多聚腺苷酸化位点，然后转录物被剪切并产生多聚腺苷酸化的 3′端。

多聚腺苷酸化信号

如果多聚腺苷酸化装置不识别转录物的末端，

而只是与转录物中间的某个位点结合，进行剪接和多聚腺苷酸化，那么这个多聚腺苷酸化位点又是怎样吸引多聚腺苷酸化装置的呢？对于这个问题的回答取决于我们讨论的是真核细胞还是病毒。先来看一看哺乳动物的**多聚腺苷酸化信号**（polyadenylation signal）。截至 1981 年，分子生物学家已经测定了许多哺乳动物基因的序列，发现这些序列中最显著的共同特征是在多聚腺苷酸化位点前都含有 20～30bp 的 AATAAA 序列。在 RNA 水平上，大多数哺乳动物 mRNA 的 poly（A）上游约 20nt 处常出现 AAUAAA 序列。Molly Fitzgerald 和 Thomas Shenk 通过两种方式测试了 AAUAAA 序列的重要性。首先，他们删除了 AAUAAA 与多聚腺苷酸化位点之间的核苷酸，然后测定所得 RNA 3′端的序列，发现只是将多聚腺苷酸化位点往后平移了大致与删除序列等同长度的距离。

这一结果说明序列 AAUAAA 至少是部分信号，能够引起下游约 20nt 处的多聚腺苷酸化。如果是这样，那么删除这些序列应该导致多聚腺苷酸化的失败。他们用 S1 实验证实了这一假设。首先，构建了一个重组病毒 SV40（命名为突变体 1471），在其晚期区末端插入相距 240bp 的两个多聚腺苷酸化信号。对该病毒晚期区转录物的 3′端做 S1 实验（第 5 章）显示，有两个相距 240bp 的多聚腺苷酸化信号（图 15.15）[在该实验中可以忽略 poly（A），因为它不与探针杂交]。因此，两个多聚腺苷酸化位点都发挥了作用，暗示对第一个多聚腺苷酸化位点有一定程度的通读。接着，他们分别删除第一个 AATAAA（突变体 1474）和第二个 AATAAA（突变体 1475），再进行 S1 实验。结果显示，删除 AATAAA 之后，该位点的多聚腺苷酸化不起作用，证明 AATAAA 对于前体 mRNA 的多聚腺苷酸化是必需的。当然，我们马上就会看到，这只是哺乳动物多聚腺苷酸化信号的一部分。

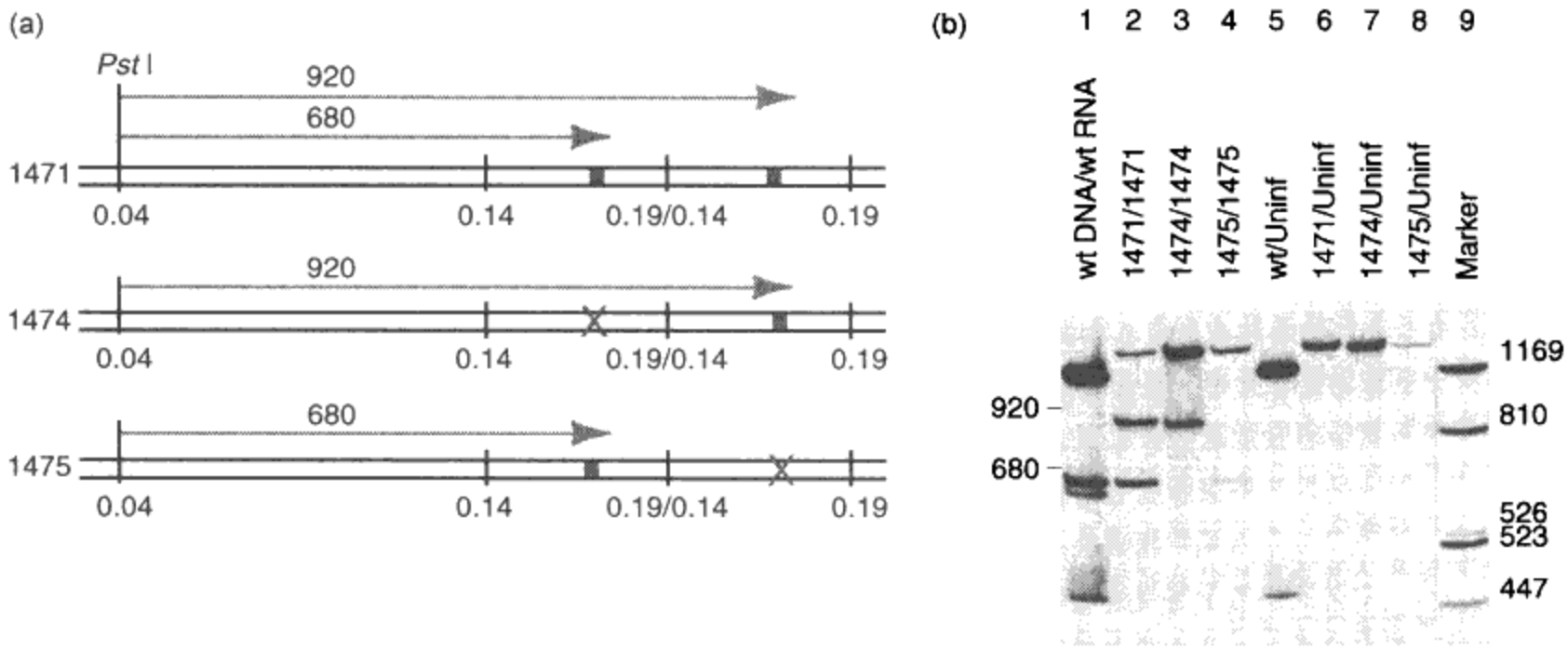


图 15.15　AAUAAA 序列在多聚腺苷酸化中的重要性。Fitzgerald 和 Shenk 构建了具有如下特征（下图）的重组病毒 SV40：（a）突变体 1471 的复制区含有相距 240bp 的两个多聚腺苷酸化位点（绿色），这两个位点从 0.14 延伸至 0.19 图距（map unit）上；突变体 1474 的上游 AATAAA 位点处的 16bp 被删除（红色），而突变体 1475 在下游 AATAAA 位点处的 16bp 被删除（红色），这两种突变导致了在前体 mRNA 中相应的 AAUAAA 序列的缺失。然后进行 S1 分析，如果上游多聚腺苷酸化信号发挥作用，所用探针可产生 680nt 的信号，如果下游多聚腺苷酸化信号发挥作用则得到 920nt 的信号（蓝色箭头）。(b) 实验结果。泳道标注：最上面是探针号，下面是 RNA（或模板）。泳道 1：野生型探针和模板，在 680nt 处显示信号，还有一个不常见的假信号；泳道 5～8 为没有感染的阴性对照。每个泳道最上面的条带是重新复性的 S1 探针，可以忽略。右图（b）是实验结果的示意图，表明删除 AAUAAA 导致此位点多聚腺苷酸化的失败。(*Source*: Adapted from Fitzgerald, M. and T. Shenk, The sequence 5′-AAUAAA-3′ forms part of the recognition site for polyadenylation of late SV40 mRNAs. *Cell* 24 (April 1981) p. 257, f. 7.)

信号序列 AAUAAA 是不变的吗？或者说它具有变异耐受性吗？早期研究中，人为设计了几种信号序列（AAUACA、AAUUAA、AACAAA 和 AAUGAA），结果表明多聚腺苷酸化不允许信号序列 AAUAAA 发生变异。但是到了 1990 年，对 269 种脊椎动物 cDNA 的多聚腺苷酸化信号的统计发现，这些天然信号序列存在一定程度的变异，特别是第二个核苷酸。Marvin Wickens 综合分析这些资料后，

证实了一段多聚腺苷酸化信号的保守序列（图15.16），即在RNA水平，最普遍的信号是AAUAAA，它启动多聚腺苷酸化最有效。最常见的变异序列是AUUAAA，启动的多聚腺苷酸化效率只有AAUAAA序列的80%。而其他变异形式更为少见，在启动多聚腺苷酸化上也更低效。

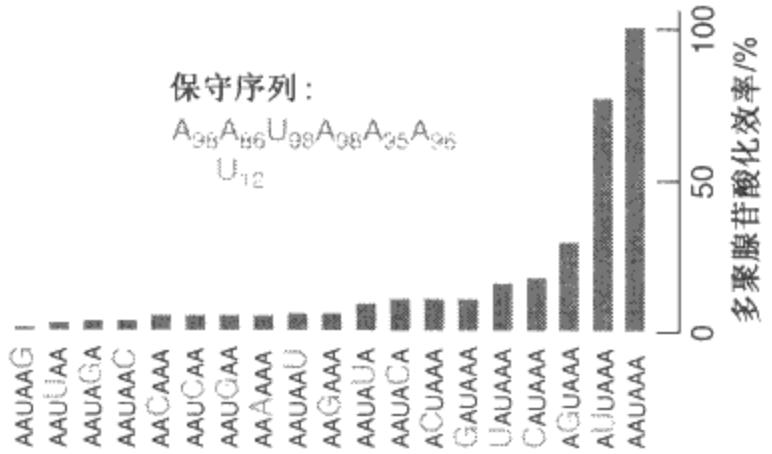


图15.16　369种脊椎动物多聚腺苷酸化数据的总结。保守序列（以RNA形式）在顶部给出，数字表示各个碱基出现的频率。第二个位置上U代替A的频率高达12%，所以单独在保守序列下面列出。接下来所列的是不同信号所引起的多聚腺苷酸化效率。与保守序列不同的碱基以大号蓝色标出。标准序列AAUAAA在最底部给出，而列在它上面的是最常见的突变体AUUAAA。（*Source*: Adapted from Wickens, M., How the messenger got its tail: addition of poly (A) in the nucleus. *Trends in Biochemical Sciences*, 15: 278, 1990.）

现在我们已经很清楚，信号序列AAUAAA本身并不足以启动多聚腺苷酸化。否则，存在于内含子中的许多AAUAAA序列下游就都会发生多聚腺苷酸化。一些研究者发现，删除紧接多聚腺苷酸化位点下游的序列可以破坏多聚腺苷酸化的发生。基于这一发现产生了如下猜测：紧接多聚腺苷酸化位点下游的序列中可能存在多聚腺苷酸化信号的另一个成分。问题是在脊椎动物中这一区间不是高度保守的，只是倾向于富含GU或U而已。

以上结果表明，有效的多聚腺苷酸化信号至少应该包括AAUAAA及其随后约20bp处富含GU或U的序列。为了验证这一假设，Anna Gil和Nicholas Proudfoot对兔β-球蛋白的高效多聚腺苷酸化信号进行了检测，这个信号含有AAUAAA序列，24bp之后是一个富GU区，之后又紧接着一个富U区。在这一部分的讨论中，我们仍将用RNA的方式来表示序列（如AAUAAA），尽管突变肯定是在DNA水平上发生的。研究者首先在天然多聚腺苷酸化信号的上游插入一个相同拷贝的多聚腺苷酸化信号，然后用S1分析法在这两个位点上检测突变体（3号克隆）的多聚腺苷酸化。结果发现突变体在新位点的多聚腺苷酸化速度是原位点的90%。证明新插入的多聚腺苷酸化位点是有活性的。接着，他们将新插入的多聚腺苷酸化信号区（位于上游）内的35bp富含GU和U的序列删除掉，构建了一个新的突变体［2（v）克隆］。结果在新位点的多聚腺苷酸化过程被破坏，再次证明这个富GU和U的35bp片段是多聚腺苷酸化信号中十分重要的一部分。

为了确定最短的有效多聚腺苷酸化位点，研究者在克隆2（v）上加上各种序列测试其多聚腺苷酸化的活性。研究发现，单独富GU或富U序列本身并不能组成一个有效的多聚腺苷酸化信号：GT克隆含富GU区，但只有野生型30%的活性。A-T克隆有富U区，也只有正常活性的30%。另外，GU/U区的位置十分重要。在C-GT/T克隆中，AATAAA片段下游的GU/U往3′端平移了16bp，结果多聚腺苷酸化活性还不到正常的10%。再则，富GU和富U区之间的间距也十分重要。GT-T克隆同时含有富GU区和富U区，只是它们之间的距离增加了5bp，结果其多聚腺苷酸化活性只有正常的30%。因此，一个有效的多聚腺苷酸化信号应该具有序列AAUAAA，接着是23～24bp之后的富GU区，再紧接其后是富U区。

植物和酵母的mRNA也进行多聚腺苷酸化，但它们的多聚腺苷酸化信号与脊椎动物的不同。酵母基因通常在多聚腺苷酸化位点附近缺少AAUAAA序列。事实上，除了在多聚腺苷酸化位点上游有一个富AU区外，很难从酵母的多聚腺苷酸化信号中得出一个固定模式。在植物基因的适当位置可能具有AAUAAA序列，删除该序列可以阻止多聚腺苷酸化。但是植物和动物的多聚腺苷酸化信号并不一样，花椰菜嵌纹病毒的单碱基替换不像在脊椎动物中会引起多聚腺苷酸化的降低。另外，把动物基因的多聚腺苷酸化信号接在植物基因的末端不能发挥作用。

小结　有效的哺乳动物多聚腺苷酸化信号应该是在其前体mRNA的多聚腺苷酸化位点上游约20nt处有一个AAUAAA序列，之后23～24bp处有一个富GU区，再紧接一个富U区。这些区域会有一定的变异，从而导致多聚腺苷酸化效率的变化。植物的多聚腺苷酸化信号通常也具有AAUAAA序列，只是这个序列在植物中比动物中有更高的变异性。酵母的多聚腺苷酸化信号更为不同，其中很少包含AAUAAA序列。

前体mRNA的剪切和多聚腺苷酸化

通常所说的多聚腺苷酸化实际上包括RNA的

剪切和多聚腺苷酸化两个过程。这一节我们先简要讨论剪切反应中所涉及的因子，然后详细讨论多聚腺苷酸化反应。

前体 mRNA 的剪切（cleavage） 在多聚腺苷酸化之前，哺乳动物前体 mRNA 的剪切需要一些蛋白质的参与。其中，最初命名为“剪切和多聚腺苷酸化因子（CPF）”，现在称为**剪切和多聚腺苷酸化特异性因子**（cleavage and polyadenylation specificity factor，**CPSF**）的蛋白质对多聚腺苷酸化是必需的。交联实验证实该蛋白质与 AAUAAA 信号结合。Shenk 及同事在 1994 年报道了另外一个参与识别多聚腺苷酸化位点的因子——**剪切激活因子**（cleavage stimulation factor，CstF）。根据交联反应结果，该因子结合到 G/U 区域。因此，CPSF 和 CstF 结合到剪切和多聚腺苷酸化位点的两侧。其中单个因子的结合不稳定，两个因子一起才可以协同稳定地结合。

剪切反应中所必需的另一对 RNA 结合蛋白是**剪切因子 I**（cleavage factor I，**CFI**）和**剪切因子 II**（**CFII**）。此外，poly（A）聚合酶本身很有可能也是剪切所必需的，因为多聚腺苷酸化紧接在剪切之后。事实上，剪切和多聚腺苷酸化之间的偶联非常紧密，几乎检测不到切下来还没有发生多聚腺苷酸化的 RNA。

另一个与剪切反应密切相关的蛋白质是 RNA 聚合酶 II。研究发现，由 RNA 聚合酶 II 体外合成的 RNA 可以被正确地加帽、剪接和多聚腺苷酸化，而由 RNA 聚合酶 I 和 III 合成的 RNA 却不行。事实上，即使由 RNA 聚合酶 II 合成的 RNA，如果其最大亚基缺少羧基末端域（carboxyl-terminal domain，CTD），这些 RNA 也不能有效地剪切和多聚腺苷酸化，表明 CTD 可能以某种方式参与 RNA 的剪切和多聚腺苷酸化。

1998 年 Yutaka 和 James Manley 证明，CTD 可激发剪切反应，而且这种激发不依赖于转录过程。首先，在其他所有剪切和多聚腺苷酸化因子存在的情况下，他们测试了磷酸化和未磷酸化的聚合酶 II（IIO 和 IIA，第 10 章）激活剪切的能力，然后分别用 CPSF、CstF、CFI、CFII、poly（A）、聚合酶 IIA 或聚合酶 IIO 孵育 ^{32}P 标记的腺病毒 L3 的 pre-mRNA。孵育之后，电泳产物，并通过放射自显影确定前体 mRNA 是否在正确位置被剪切。图 15.17 是实验结果，即聚合酶 IIA 和聚合酶 IIO 都可以激发前体 mRNA 的正确剪切，并得到预期大小的 5′和 3′的片段。

为了验证聚合酶 II 的 CTD 是激发剪切的重要成分，Hirose 和 Manley 把 CTD 与 GST 融合表达（第 4 章），然后用 GST 亲和层析纯化融合蛋白。对

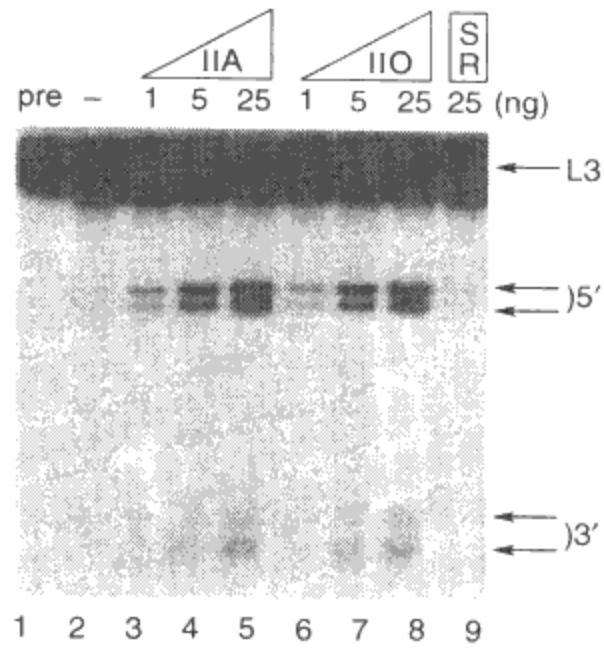


图 15.17 RNA 聚合酶 IIA 和 IIO 对未多聚腺苷酸化 mRNA 体外剪切的影响。Hirose 和 Manley 制备了一个 ^{32}P 标记的腺病毒 L3 前体 mRNA，并把它与所有剪切和多聚腺苷酸化因子［CPSF、CstF、CFI、CFII 和 poly（A）聚合酶］一起孵育，并在其中分别加上如图所示的 RNA 聚合酶 IIA、IIO、无蛋白（–）、纯化的 HeLa 细胞 SR 蛋白（各种蛋白质的量是纳克级的），然后电泳 RNA 产物并放射自显影检测。结果显示，前体 mRNA 的 3′和 5′剪切位点都是正确的，RNA 聚合酶 IIA 和 IIO 都可以激发 pre-mRNA 3′端和 5′端的正确剪切。（*Source*：Hirose，Y. and Manley，J. RNA polymerase II is an essential mRNA polyadenylation factor. *Nature* 395（3 Sep 1998）f. 2，p. 944. Copyright © Macmillan Magazines Ltd.）

部分融合蛋白的 CTD 进行磷酸化处理，并在腺病毒 L3 的前体 mRNA 剪切实验中测定磷酸化和未磷酸化融合蛋白的作用。图 15.18（右）中显示磷酸化和未磷酸化的 CTD 都可以激发剪切，但是磷酸化 CTD 的作用比未磷酸化的约强 5 倍。这是合理的，因为在聚合酶 IIO 中 CTD 是以磷酸化形式存在并实施转录的。但在图 15.17 中，用完整 RNA 聚合酶 II 进行反应，CTD 有无磷酸化对多聚腺苷酸化没有影响，其原因还不清楚。

如果 CTD 是激发前体 mRNA 剪切的主要因素，那么聚合酶 IIB（聚合酶 IIA 的水解产物）在缺少 CTD 的情况下应该不能激发剪切反应。图 15.18（左）中的结果确实如此。因此，RNA 聚合酶 II，尤其是 CTD，看来是多聚腺苷酸化之前前体 mRNA 有效剪切所必需的。图 15.19 显示剪切前结合到前体 mRNA 上的蛋白质复合体。

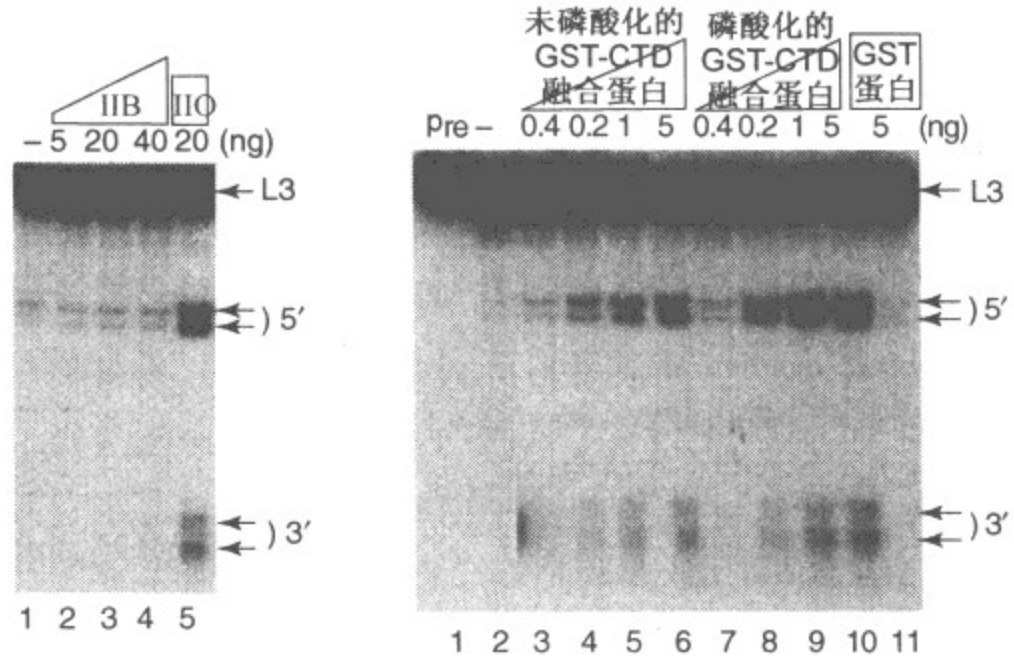


图 15.18　RpB1CTD 对前多聚腺苷酸化 mRNA 体外剪切的影响。Hirose 和 Manley 把标记的前体 mRNA 与剪切和多聚腺苷酸化因子一起孵育，并按图 15.17 测定剪切反应。如右图所示，在剪切反应中加入磷酸化的、未磷酸化的 GST-CTD 融合蛋白或单独的 GST。如左图所示，在剪切反应中加入 RNA 聚合酶 IIB 或 IIO。磷酸化 CTD 比未磷酸化 CTD 能更有效地激发剪切反应。缺乏 CTD 的 RNA 聚合酶 IIB 不能激发多聚腺苷酸化反应。(*Source*: Hirose, Y. and Manley, J. RNA polymerase II is an essential mRNA polyadenylation factor. *Nature* 395 (3 Sep 1998) f. 3, p. 94. Copyright © Macmillan Magazines Ltd.)

剪切和多聚腺苷酸化特异因子

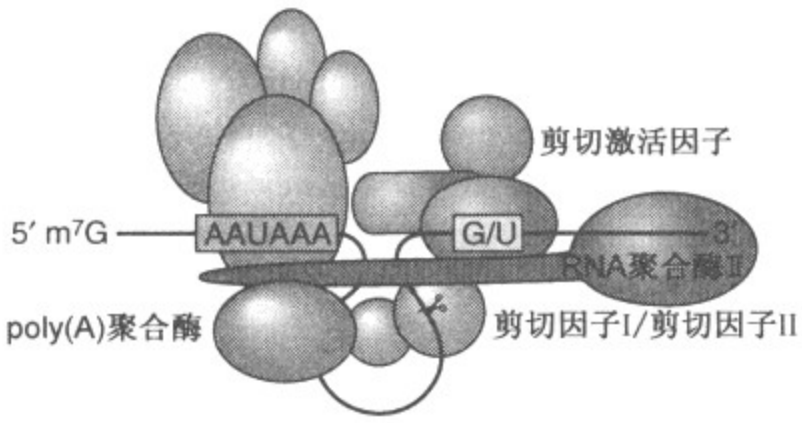


图 15.19　剪切复合体前体模型。模型展示了所有可能参与剪切的蛋白质的位置以及多聚腺苷酸化信号的两个部分（绿色和黄色）的相对位置。(*Source*: Adapted from Wahle, E. and W. Keller, The biochemistry of polyadenylation, *Trends in Biochemical Sciences* 21 [1996] pp. 247-250, 1996.)

小结　多聚腺苷酸化的发生需要前体 mRNA 的剪切和在剪切处的多聚腺苷酸化。在哺乳动物中剪切反应的发生需要 CPSF、CstF、CFI、CFII、poly（A）聚合酶和 RNA 聚合酶 II（尤其是 Rpb1 亚基中的 CTD）。

多聚腺苷酸化反应的起始　一旦前体 mRNA 在其 AAUAAA 基序下游被剪切后，就可以分两个阶段进行多聚腺苷酸化。首先是起始，该阶段依赖于信号序列 AAUAAA，给前体 mRNA 缓慢加上至少 10 个 A。第二个阶段是延伸，该阶段不依赖信号序列 AAUAAA，而是依赖第一个阶段加上去的寡聚 A。该阶段迅速地在 RNA 尾部加上 200 个甚至更多的 A。我们先从起始阶段开始介绍。

严格地讲，所谓的“多聚腺苷酸化信号”实际上是一种剪切信号，吸引剪切酶在 AAUAAA 下游约 20nt 处剪切 RNA。而多聚腺苷酸化，就是在剪切酶产生的 3′端添加上寡聚 A，其本身并不能识别剪切信号。这是因为剪切酶已经在多聚腺苷酸化之前切去了下游的部分信号（富 GU 区和富 U 区）。

是什么信号引起了多聚腺苷酸化反应呢？答案似乎是 AAUAAA 及紧跟其后的至少 8nt 的核苷酸。之所以得出这个结论是因为人工合成的含有 AAUAAA 的寡聚核苷酸（只有 11nt）在体外能够被多聚腺苷酸化，并且 RNA 末端与 AAUAAA 之间的最适长度是 8nt。

为了在体外研究多聚腺苷酸化过程，有必要将多聚腺苷酸化从剪切反应中分离出来。分子生物学家用带有标记的、短的，且在离 3′端至少 8nt 处有 AAUAAA 序列的 RNA 实现了这一点。这些底物可以模拟已经被剪切的、将要被多聚腺苷酸化的前体 mRNA。用于检测多聚腺苷酸化的实验是电泳分析标记的 RNA。如果 RNA 已经加上了 poly（A），它就会变大，电泳时会移动得缓慢。各种不同大小的片段也会被连续地分离开来，因为 poly（A）尾的长度在不同的分子内有所不同。在这一节里，我们

将用“多聚腺苷酸化”这一术语表示在这种模型底物RNA的3′端添加poly（A）。

图15.20展示了Marvin Wickens及同事如何通过实验来研究多聚腺苷酸化所必需的两个成分：poly（A）聚合酶和特异性因子。现在我们已经知道这个特异性因子就是CPSF。在底物浓度较高时，poly（A）聚合酶催化poly（A）添加到任何RNA的3′端，而当底物浓度较低时，poly（A）聚合酶就不能进行多聚腺苷酸化（泳道1）。可以识别AAUAAA信号的CPSF也不能独自完成加尾（泳道2）。但是两者一起加入，就能使底物多聚腺苷酸化（泳道3）。泳道4中显示存在异常信号序列AAUCAA时，混合物不能引起多聚腺苷酸化反应。

他们还用不同的模型RNA作为加尾底物进行实验，证明多聚腺苷酸化本身包含两个阶段。第一种RNA是SV40晚期mRNA末端的58nt序列，其中含有AAUAAA信号，与图15.20中用到的相同。第二种RNA是在第一种RNA的3′端加了40个A[短poly（A）]。第三种RNA是在第一种RNA的3′端加了载体分子上的40个核苷酸，而不是短poly（A）。同时，他们还合成了另一组的3个底物，其中序列AAUAAA被AAGAAA代替。

Sheets和Wickens用各种底物与HeLa细胞的细胞核提取物进行标准的多聚腺苷酸化反应。在图15.21中，泳道1～4显示提取物能够使正常的带有AAUAAA信号的模型底物多聚腺苷酸化。泳道5～8显示末端有40个A（A_{40}）的模型底物也可以被多聚腺苷酸化，但底物的放射性强度较弱，且提取物不能将未含40个A的底物（X_{40}）多聚腺苷酸化。泳道13～16证明提取物不能使含异常信号AAGAAA且无poly（A）的底物多聚腺苷酸化。泳道17～20显示了最重要的一点，提取物能够使含有异常信号AAGAAA且在末端有40个A的底物多聚腺苷酸化。因此，只要加上40个A，多聚腺苷酸化就不再依赖于AAUAAA信号，但是这些外加的核苷酸必须是多聚腺苷酸。泳道21～24显示带有一个异常信号的X_{40}底物不能被多聚腺苷酸化。

他们进一步研究发现，能够抑制AAUAAA突变效应的最短poly（A）为9个A，但是10个A作用会更好。由此得出如下假设：在前体mRNA剪切后，多聚腺苷酸化反应依赖于AAUAAA信号和CPSF，直到poly（A）达到约10个A，随后的多聚腺苷酸化不再依赖于AAUAAA和CPSF，而是依赖于RNA 3′端的poly（A）尾。

如果CPSF能够识别多聚腺苷酸化信号AAUAAA，那么CPSF应该结合到前体mRNA的信号序列上。Walter Keller小组通过凝胶阻滞实验和RNA-蛋白质交联实验直接证实了这一点。图15.22是两个实验的结果。图15.22a是CPSF结合到被标记的、含AAUAAA信号序列的RNA上，但是对同样的RNA，如果信号序列AAUAAA中的U突变为G，则不能和CPSF结合。图15.22b表明含有序列AAUAAA的寡聚核苷酸链可与CPSF制备液中的两个多肽进行（35kDa和160kDa）交联。当含有未标记的、带有AAUAAA信号的竞争RNA时，这些复合物不能形成。带有AAGAAA信号的竞争RNA没有这种竞争能力。所有结果都支持CPSF直接结合到AAUAAA基序上这一结论。

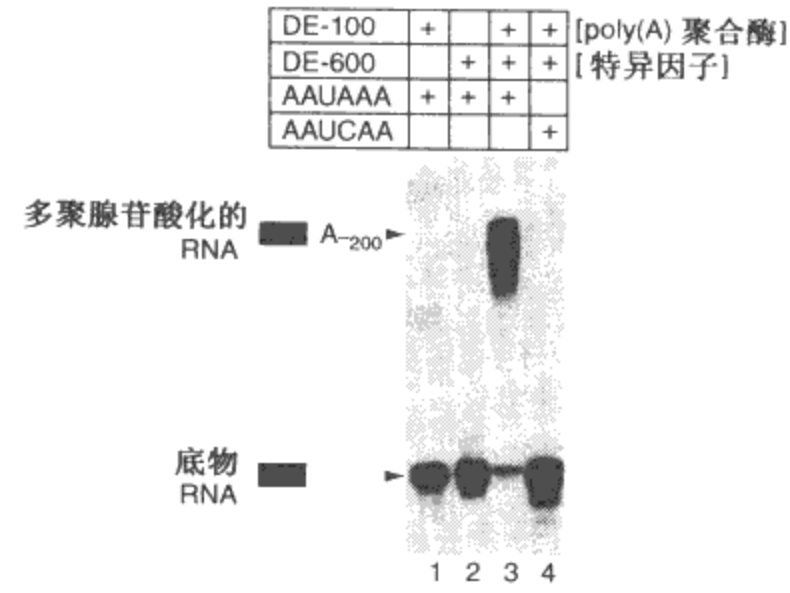


图15.20　poly（A）聚合酶和特异性因子的分离。Wichens小组通过DEAE-琼脂糖层析法分离了HeLa细胞的poly（A）聚合酶和特异性因子。该聚合酶在100mmol/L的盐溶液中稀释，所以称为DE-100组分。特异性因子在600mmol/L的盐溶液中稀释，所以称为DE-600组分。通过人工合成带有标记的、SV40晚期mRNA的－58到＋1核苷酸底物，测定分离活性。把这两部分一起或分开与底物及ATP孵育后，进行电泳，放射自显影。每个泳道的组分在图上方列出，底物的位置和多聚腺苷酸化产物在图左侧标出。（*Source*：Bardwell，V. J.，D. Zarkower，M. Edmonds，and M. Wickens，The enzyme that adds poly（A）to mRNAs is a classical poly（A）polymerase. *Molecular and Cellular Biology* 10（Fed 1990）p. 847，f. 1. American Society for Microbiology.）

小结　与新生mRNA 3′端类似的短RNA能够被多聚腺苷酸化。剪切底物的最佳多聚腺苷酸化起始信号是AAUAAA及其后连接的至少8个核苷酸。一旦poly（A）达到10nt，接下来的多聚腺苷酸化将不再依赖于信号AAUAAA，而是依赖于poly（A）本身。参与多聚腺苷酸化起始

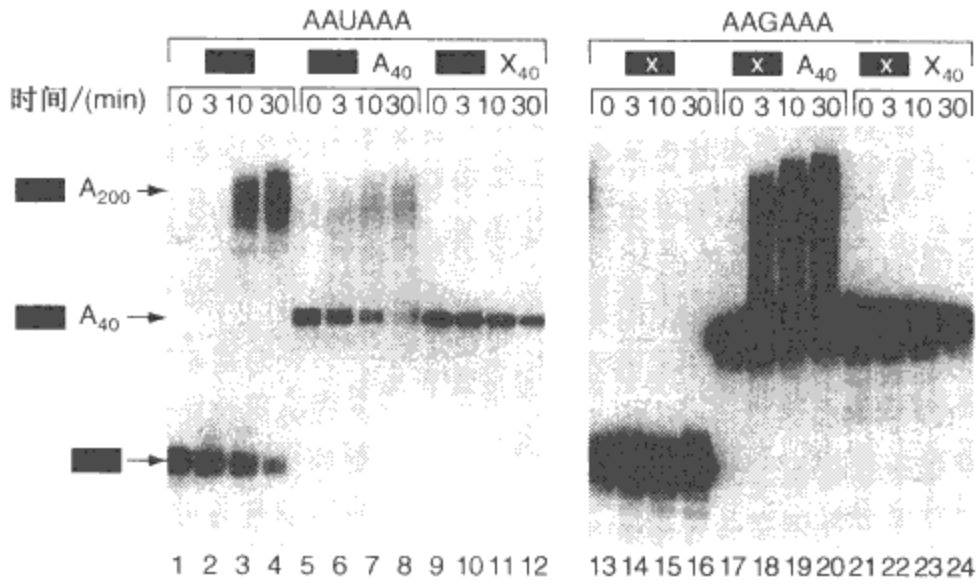


图 15.21　多聚腺苷酸化分两个阶段的证明。Sheets 和 Wickens 以 HeLa 细胞核提取物用下列标记的底物进行多聚腺苷酸化反应。①含有 SV40 晚期 mRNA 3′端序列的标准 58nt 底物，用黑框表示；②具有 40nt poly（A）的同样 RNA，黑框 A_{40} 表示；③在 3′端有 40nt 载体序列的同样 RNA，黑框 X_{40} 表示。在底物 1～3 中异常信号序列 AAGAAA 代替 AAUAAA，用黑框标白色 X 表示。对每种底物有 4 个不同的反应时间，每个泳道中所含底物在泳道上方标出，底物和产物的迁移位置在左边标出。(*Source*: Sheets and Wickens, Two phases in the addition of a poly (A) tail. *Genes & Development* 3 (1989) p. 1402, f. 1. Cold Spring Harbor Laboratory Press.)

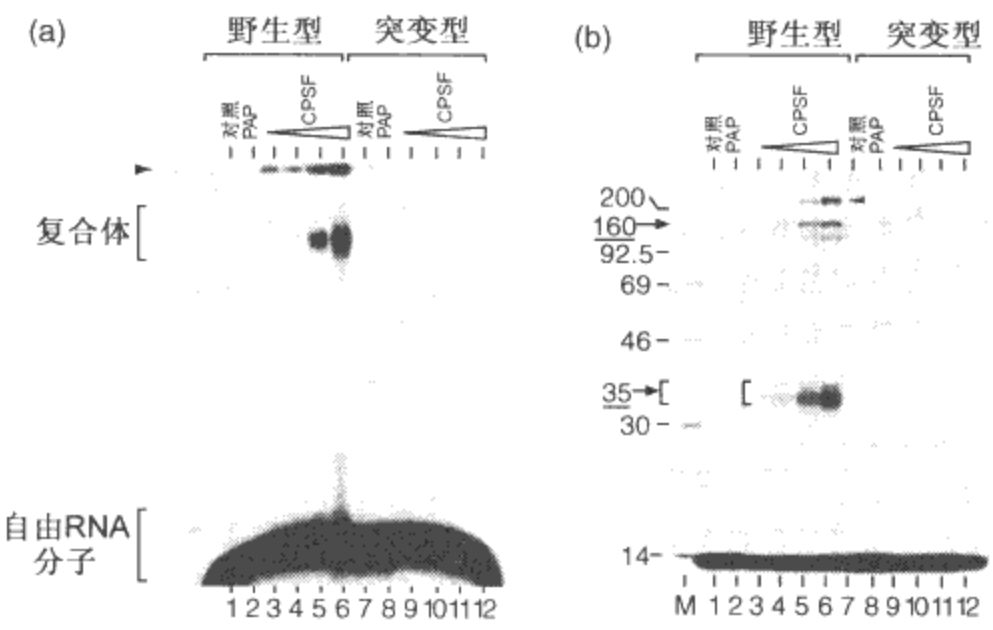


图 15.22　CPSF 结合到 AAUAAA 基序上。(a) 凝胶阻滞实验。Keller 等把标记的寡聚核苷酸链与 poly（A）聚合酶（PAP）和不同浓度的 CPSF 混合。野生型寡聚核苷酸链中含有 AAUAAA 序列，突变型寡聚核苷酸链中含有 AAGAAA 序列，对照中不含有任何外加蛋白质。CPSF 可以和野生型形成复合物但不能和突变型形成复合物。图上部箭头所示条带为残留在胶上部的物质，而非特异性条带。(b) 蛋白质和寡聚核苷链交联反应的 SDS-PAGE 分析。用紫外光照射显示（a）中蛋白质与寡聚核苷酸的交联，然后在 SDS-PAGE 胶上电泳，大条带分别出现在 35kDa 和 160kDa 处。(*Source*: Keller, W., S. Bienroth, K. M. Lang, and G. Christofori, Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3′ processing signal AAUAAA. *EMBO Journal* 10 (1991) p. 4243, f. 2.)

过程的两个蛋白质是 poly（A）聚合酶和结合到 AAUAAA 基序上的 CPSF。

poly（A）的延伸　我们知道，10nt 或更长已起始 poly（A）链的延伸不再依赖于 CPSF。然而，纯化的 poly（A）聚合酶本身并不能高效地结合并延伸 poly（A）。这暗示还有另外一个特异性因子，能够识别已起始合成的 poly（A）序列并指导 poly（A）聚合酶延伸。Elmar Wahle 纯化了一个具有这种特征的 poly（A）结合蛋白。

图 15.23b 是纯化 poly（A）结合蛋白最后一步

所得组分的 PAGE 结果，图中可以看到一个较大的 49kDa 的多肽和一个较小的多肽。由于后一条带的丰度在不同样品中不同，在一些制备液中甚至看不到，Wahle 认为这个小分子多肽不是 poly（A）结合蛋白。通过硝酸纤维素滤膜结合实验测定 49kDa 部分的 poly（A）结合能力（图 15.23a），结果显示其 poly（A）结合活性与其含量正相关。接着他又测试了 49kDa 多肽在 poly（A）聚合酶和 CPSF 存在时对多聚腺苷酸化反应的激活能力。结果再次显示其 poly（A）结合活性与 49kDa 多肽的含量正相关。因此，49kDa 多肽是一种 poly（A）结合蛋白，与此前在细胞质中发现的 70kDa poly（A）结合蛋白（**PAB I**）不一样，Wahle 将其命名为 **poly（A）结合蛋白 II**［poly（A）-binding protein II，**PABII**］。

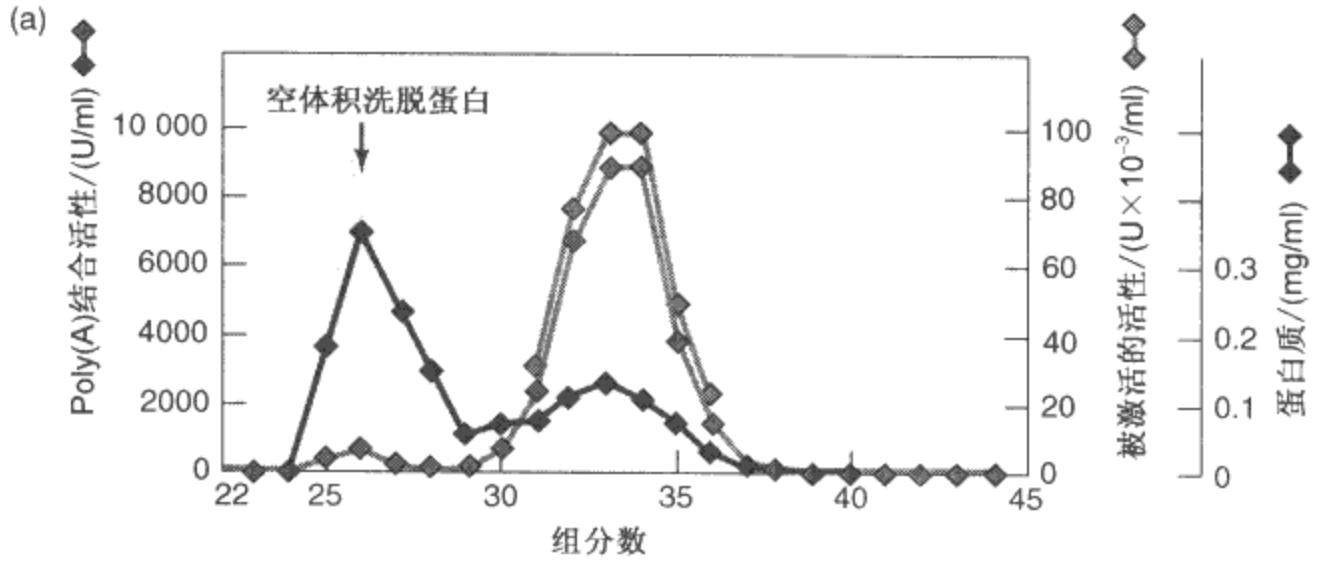


图 15.23　poly（A）结合蛋白的纯化。（a）结果总结。Wahle 利用 Sephadex G-100 对 poly（A）结合蛋白进行凝胶过滤层析纯化。用 3 个参数分别对 G-100 柱的分离组分数作图。红色代表由过滤结合实验得出的 poly（A）结合活性。绿色代表多聚腺苷酸化的激活活性（图 c）。蓝色代表蛋白质浓度。“void”表示空柱体积洗脱蛋白，这些大蛋白质不能进入凝胶颗粒的孔隙中。（b）SDS-PAGE 分析。把（a）中 G-100 柱得到的蛋白质进行 SDS-PAGE 并用考马斯亮蓝染色，实验中所用标准分子的大小在左边标出。49kDa 多肽（32～35）在 poly（A）结合活性和激发多聚腺苷酸化活性达到最高点时，其浓度也达到最大。（c）多聚腺苷酸化激活活性实验。把 G-100 柱上收集的每一部分加到标记 L3 前体 RNA 的标准多聚腺苷酸化反应体系中。泳道 1 仅含底物不含 poly（A）聚合酶，poly（A）大小的增加显示了 poly（A）结合蛋白的激发活性，这一活性在 32～35 间达到最高。（*Source*：Wahle，E.，A novel poly（A）-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation. *Cell* 66（23 Aug 1991）p. 761，f. 1. Reprinted by permission of Elsevier Science.）

像 CPSF 一样，PABII 也能激发模型底物的多聚腺苷酸化，但它结合 poly（A）而不是 AAUAAA 信号序列。说明 PABII 是在多聚腺苷酸化的延伸阶段，而不是起始阶段起作用。如果真是这样，那么其底物偏爱性就应该与 CPSF 不同。具体说，它只激发那些已经有 poly（A）尾的 RNA 的多聚腺苷酸

化，而不激发那些不含 poly（A）尾的 RNA。图 15.24 所示的结果证实了这一点。图 15.24a 显示无 poly（A）的 RNA（L3 的前体 mRNA）能够被 poly（A）聚合酶（PAP）和 CPSF 多聚腺苷酸化，但不能被 PAP 和 PABII 多聚腺苷酸化。而 PAP 与 CPSF、PABII 一起能使这个底物多聚腺苷酸化得最好。可以推测，CPSF 作为多聚腺苷酸化反应的起始因子，一旦底物被加上一段 poly（A），PABII 接着就能指导底物的多聚腺苷酸化，并且比 CPSF 做得更好。可以预测，如果 L3 前体底物的信号序列 AAUAAA 突变为 AAGAAA，则任何多聚腺苷酸化因子的组合都不能使其多聚腺苷酸化，因为多聚腺苷酸化的起始依赖于 CPSF，而 CPSF 依赖于 AAUAAA 序列。

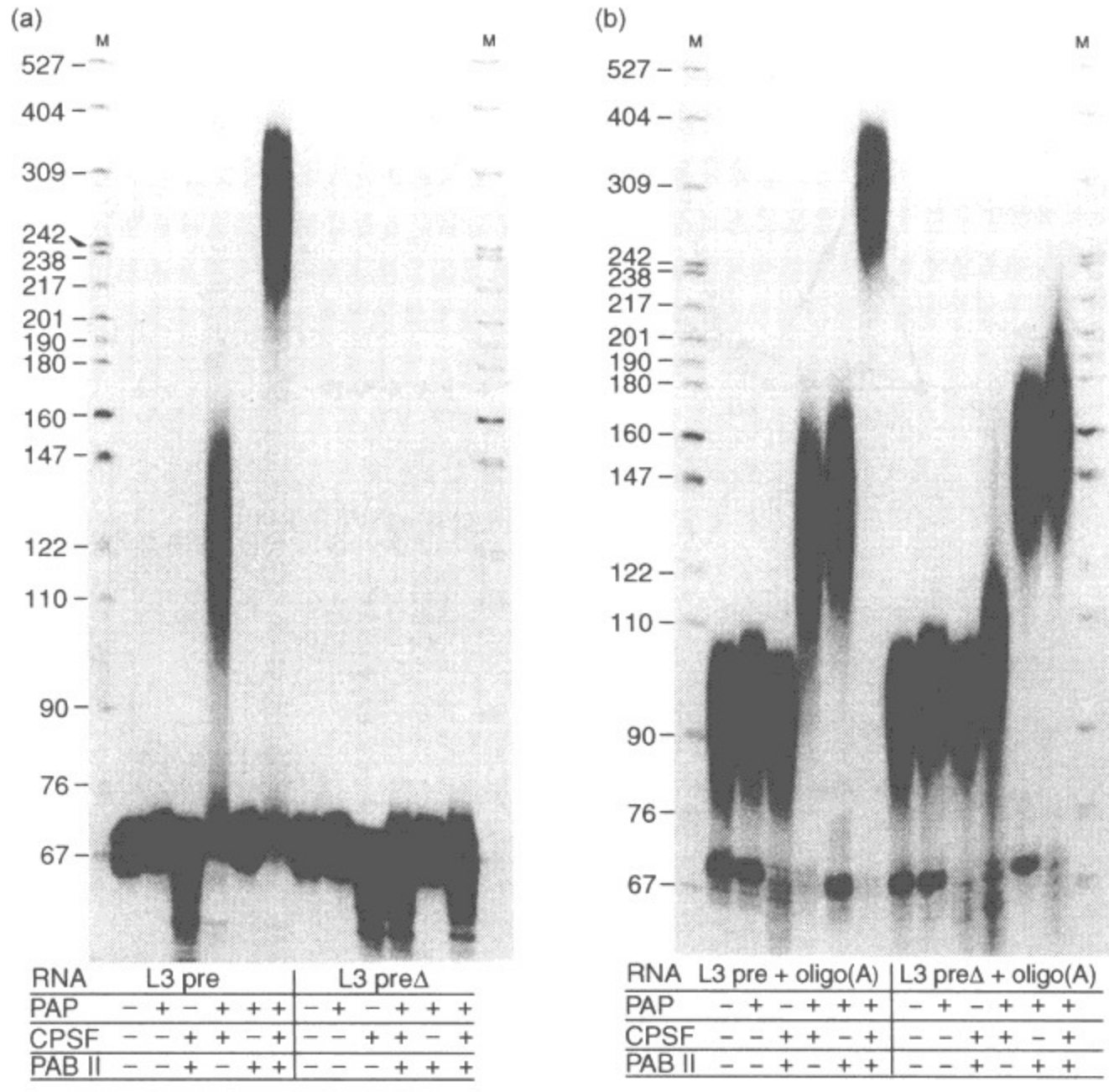


图 15.24　CPSF 和 PABII 对模型底物多聚腺苷酸化的影响。（a）缺乏 poly（A）的 RNA 的多聚腺苷酸化。在 RNA 和底部所示蛋白质存在时，进行多聚腺苷酸化反应。L3-pre 是带有信号序列 AAUAAA 的标准底物 RNA。L3-preΔ 中除了 AAUAAA 突变为 AAGAAA 外，其他都与 L3-pre 一样。没有 CPSF 帮助，PABII 不能指导 L3 pre 的多聚腺苷酸化。(b) 含有寡聚 A 的 RNA 的多聚腺苷酸化反应。除了底物的 3′端多了寡聚 A 外，其他条件都与（a）一样，这就允许 PABII 能够在缺乏 CPSF 及底物信号 AAUAAA 突变情况下仍然能起作用。两个图的左边第一道和右边最后一道都是标准分子。(*Source*：Wahle，E.，A novel poly（A）-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation. *Cell* 66（23 Aug 1991）p. 764，f. 5. Reprinted by permission of Elsevier Science.)

图 15.24b 显示同样带有 poly（A）尾的 RNA 在不同反应体系下的不同行为。它能够在 PAP 与 CPSF 或 PABII 组合时被 PAP 多聚腺苷酸化。这种现象是合理的，因为该底物带有能被 PABII 识别的 poly（A）。令人感兴趣的是，当这两个因子结合在一起时会使底物更好地多聚腺苷酸化。这也提示了在延伸阶段 PAP 能够直接或间接地和这两个因子相互作用。图 15.24b 证实在缺乏 CPSF 时，只要一开始 RNA 带有寡聚 A，PABII 就能够指导带有突变 AAGAAA 序列的 RNA 进行有效的多聚腺苷酸化。

这种情况也是合理的，因为寡聚 A 为 PABII 提供了识别位点，使其不再依赖 CPSF 和 AAUAAA 信号序列。图 15.25 是一个多聚腺苷酸化起始和延伸的模型。在起始阶段，最佳活性需要 PAP、CPSF、CstF、CFI、CFII 和两部分多聚腺苷酸化信号，即多聚腺苷酸化位点两侧的 AAUAAA 和 G/U 富集区，而在多聚腺苷酸化的延伸阶段需要 PAP、PABII 和至少 10nt 长的 poly（A）。这一阶段还可以被 CPSF 加强。表 15.2 列出了所有这些蛋白质因子及其结构与功能。

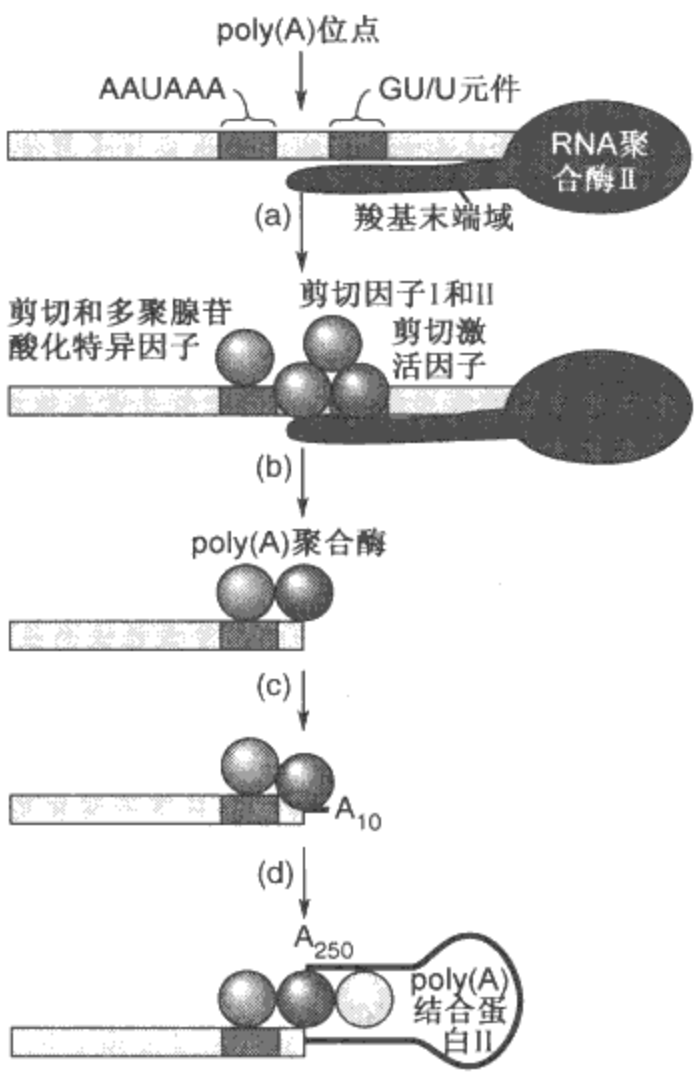


图 15.25　多聚腺苷酸化反应模型。（a）CPSF（蓝色）、CstF（褐色）、CFI（灰色）和 CFII（灰色）在信号序列 AAUAAA 和 GU/U 富集区的指导下组装在 pre-mRNA 上；（b）在 RNA 聚合酶 II 的 CTP 激发下发生剪切，接着，CPSF、CFI 和 CFII 离开复合体，同时 poly（A）聚合酶进入；（c）在 CPSF 协助下，poly（A）聚合酶（橙色）起始 poly（A）的合成，最后得到至少 10nt 的寡聚 A。（d）PABII（黄色）进入复合体并延长寡聚 A 至全长 poly（A）。在这一点上复合物可能解体。

表 15.2　哺乳动物细胞 3′剪切和多聚腺苷酸化所需的因子

因子	多肽/kDa	特点
poly（A）聚合酶（PAP）	82	参与剪切和多聚腺苷酸化，催化 poly（A）合成
剪接和多聚腺苷酸化特异因子（CPSF）	160	参与剪切和多聚腺苷酸化，结合 AAUAAA 并与 PAP 和 CstF 相互作用
	100	
	70	
	30	
剪切激活因子（CstF）	77	仅参与剪切，与下游元件结合并与 CPSF 相互作用
	64	
	50	
剪切因子 I（CFI）	68	仅参与剪切，与 RNA 结合
	59	
	25	
剪切因子 II（CFII）	未知	仅参与剪切
RNA 聚合酶 II（特别是 CTD）	许多	仅参与剪切
poly（A）结合蛋白 II（PABII）	49	激发 poly（A）的延伸，与延伸的 poly（A）尾结合，参与 poly（A）尾的延伸

Source: Reprinted from Wahle, E. and W. Keller, The biochemistry of polyadenylation, *Trends in Biochemical Sciences* 21: 247-250. Copyright © 1996 with permission of Elseiver Science.

小结　对哺乳动物而言，poly（A）的延伸需要一个称为 poly（A）结合蛋白 II（PABII）的特殊因子。这个蛋白质结合到已经起始合成的 poly（A）上，并通过 poly（A）聚合酶把 poly（A）延长到 250nt 或更长。PABII 独立于信号序列 AAUAAA 而起作用，它仅依赖于 poly（A），但其活性可被 CPSF 加强。

多聚腺苷酸酶

1991 年，James Manley 及同事克隆了编码牛多聚腺苷酸酶的 cDNA。测序结果表明，两种不同的 cDNA 由于选择性剪接而具有不同的 3′端，从而产生

了C端不同的两种PAP（PAPI和PAPII）。PAPII蛋白有几个区域的序列或多或少地与其他蛋白质已知的功能域匹配。按照从N端到C端的顺序，这些区域分别是：1个RNA结合域（RBD）、1个多聚酶模块（polymerase module，PM）、2个核定位信号（NLS1和NLS 2）、几个丝氨酸/苏氨酸富含区（S/T）。到1996年，又发现了另外4种PAP的cDNA。其中2个比较短，可能是由前体mRNA多腺苷酸化形成的。另一个比较长，可能来源于一个假基因。在大多数组织中最重要的PAP可能是PAPII。

由于推测催化多腺苷酸化反应的多聚酶模块位于蛋白质的氨基末端，那么至少需要多长的羧基端才能保持酶活性就成为一个有意思的问题。为了检测羧基端的重要性，James Manley及同事通过SP6 RNA聚合酶体外转录了全长的或3′端不同程度缺失的PAP的cDNA，然后在无细胞网织红血球中翻译表达，产生了一个689个氨基酸的全长蛋白质和缩短的分别含538、379和308个氨基酸的蛋白质。然后检测这些蛋白质在有牛胸腺CPSF时的特异性多聚腺苷酸化活性。全长序列和含538个氨基酸的蛋白质具有活性，而较小的蛋白质没有。因此，S/T域不是活性所必需的，但是从聚合酶模块向C端延伸至少150个氨基酸对于保证体外条件下酶的活性是必需的。

小结 对编码牛胸腺多聚腺苷酸酶的cDNA进行克隆和测序的结果表明，其中含有5种cDNA的混合物，分别由不同的选择性剪接和不同的多聚腺苷酸化产生。根据最长序列预测的酶结构包括一个RNA结合区（RBD）、一个多聚酶模块（PM）、两个核定位信号（NLS1和2）和几个丝氨酸/苏氨酸富含区（S/T）。除最后一个区域，其他都是体外酶活性所必需的。

poly(A)的更新

图15.7显示了细胞核poly（A）和细胞质poly（A）在大小上略有差异。该实验用的是新标记的RNA，所以poly（A）没有足够的时间被降解。Sheiness和Darnell对细胞中被连续标记48h的前体mRNA进行了另外的研究，获得了一批分子大小“稳定”的poly（A），即在任何给定时间内都可以观察到细胞内分子质量自然大小的poly（A）。图15.26显示细胞核poly（A）和细胞质poly（A）在大小上具有明显差异。核poly（A）的主峰是（210±20）nt，而细胞质poly（A）的主峰是（190±20）nt。此外，细胞质poly（A）的峰在趋向小片段的区域比核poly（A）的更宽。这种宽峰至少包括50nt。可见，poly（A）在细胞质中似乎被截短了。

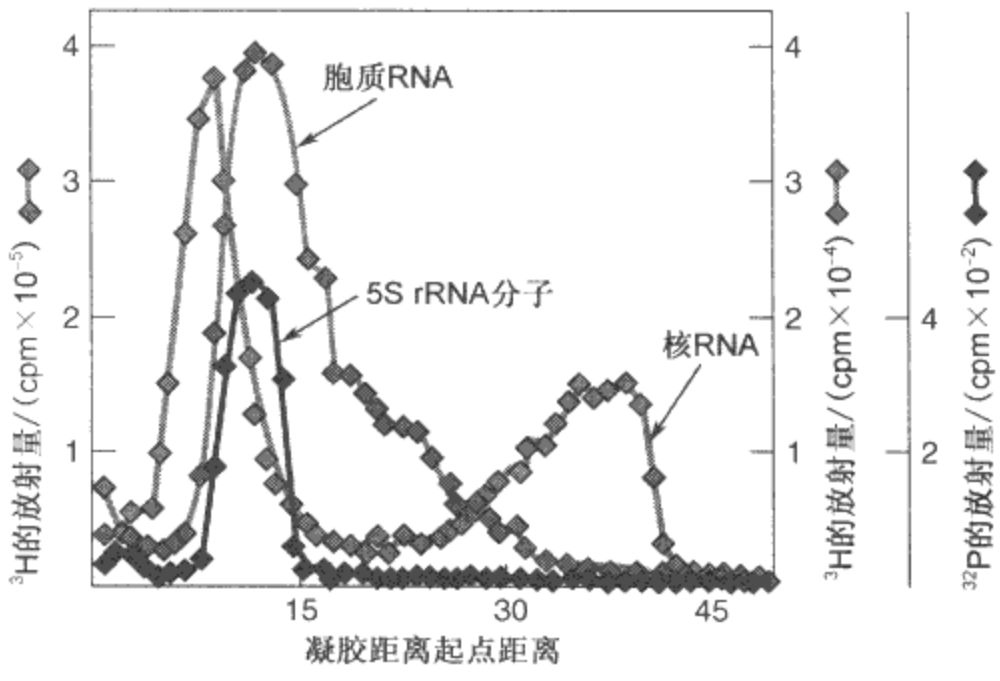


图15.26 细胞质多聚腺嘌呤的缩短。Sheiness和Darnell用^{3}H-腺嘌呤标记HeLa细胞48h，然后分离核（绿色）和细胞质（红色）poly（A）$^+$RNA，凝胶电泳分析，电泳过程中用含［^{32}P］的5S rRNA（蓝色）作为分子质量标准。(*Source*：Adapted from Sheiness，D. and J. E. Darnell，Polyadenylic acid segment in mRNA becomes shorter with age. *Nature New Biology* 241：266，1973.)

1970年，Maurice Sussman提出了“入场券”假说，认为每个mRNA都有一张“入场券”被允许进入核糖体进行翻译，mRNA每翻译一次，都会被打上一个“票孔”。当mRNA积累了足够多的“票孔”时，便不再被翻译了。poly（A）是理想的“入场券”，每翻译一次，poly（A）的长度就会缩短一些。为了验证这一观点，Sheiness和Darnell检测了正常状态以及有吐根碱（emetine）（抑制翻译）存在时，胞质poly（A）缩短的速率。结果发现，无论翻译发生与否，胞质poly（A）的大小并无差异。

因此，poly（A）的缩短并不是依赖于翻译的进行，如果存在“入场券”的话，似乎也不是 poly（A）。

poly（A）在细胞质中也并不总是缩短，它还会更新。也就是说，poly（A）不断地被 RNase 降解，又被胞质 poly（A）聚合酶延长。但总趋势是缩短，最终，mRNA 将会失去几乎所有的 poly（A）。那时，mRNA 的寿命也就接近终结了。

小结　poly（A）在细胞质中反复地缩短又延长。RNA 酶使其缩短，poly（A）聚合酶再让它延长。当 poly（A）耗尽时，mRNA 也就面临降解了。

细胞质内的多聚腺苷酸化　对细胞质内多聚腺苷酸化研究得最为深入的是卵母细胞成熟过程中的例子。例如，体外非洲爪蟾卵母细胞的成熟是由孕酮刺激的，未成熟卵母细胞质中储存了大量的 mRNA，被称为**母系信息**（maternal message）或**母系 mRNA**（maternal mRNA），其中有许多几乎是完全去腺苷酸化的（deadenylated），并且也没有翻译成蛋白质。在卵母细胞成熟过程中，一些母系 mRNA 被多聚腺苷酸化，另一些则被去腺苷酸化。

为了找出控制成熟卵细胞内特异的细胞质多聚腺苷酸化的关键因子，Wickens 及同事将以下两种 mRNA 分别注入爪蟾的卵母细胞质中。第一种是合成的 D7 mRNA 的 3′端片段，这是非洲爪蟾的一个 mRNA，已知会经历与成熟有关的多聚腺苷酸化。第二个是合成的 SV40 mRNA 的 3′端片段。如图 15.27 所示，D7 mRNA 是多聚腺苷酸化的，而 SV40 mRNA 却没有，表明 D7 mRNA 含有一个或多个序列是成熟卵母细胞特有的多聚腺苷酸化所必需的，而 SV40 mRNA 缺少这种序列。

Wickens 及同事注意到，在卵母细胞成熟过程中经历多聚腺苷酸化的爪蟾 mRNA 都具有 UUUUUAU 序列，或者一个与之密切相关的上游信号序列 AAUAAA。这是关键所在吗？为了找出答案，他

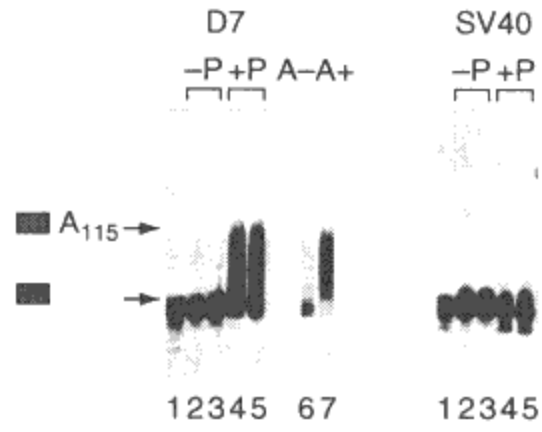


图 15.27　两种 RNA 的多聚腺苷酸化与细胞成熟度的关系。Wickens 及同事将标记的 RNA 注射进入非洲爪蟾卵母细胞，并由孕酮激发其成熟特异性多聚腺苷酸化，培养 12h 后，分离标记的 RNA 产物，电泳，放射自显影并观察。两种 RNA 一种是合成的非洲爪蟾 mRNA 的 3′端片段，通常要进行成熟特异性多聚腺苷酸化，另一种是 SV40 mRNA，不进行成熟特异性多聚腺苷酸化，未多聚腺苷酸化的 RNA 和带有 115nt poly（A）的 RNA 的迁移位置在左边用红框标记出。培养过程中有和无孕酮分别用＋P 和－P 标记，泳道 6 和 7 含有经寡聚胸腺嘧啶-纤维素层析的 RNA。没有结合到树脂上的 RNA 标为 A－，结合到树脂上的标为 A＋。（*Source*：Fox et al.，Poly（A）addition during maturation of frog oocytes：Distinct nuclear and cytoplasmic activities and regulation by the sequence UUUUUAU. *Genes & Development* 3（1989）p. 2154. f. 3. Cold Spring Harbor Laboratory Press.）

们将 UUUUUAU 插入 SV40 mRNA 的 AAUAAA 的上游并重新检测。图 15.28 证明该序列的插入使 SV40 RNA 多聚腺苷酸化了。据此，UUUUUAU 序列被命名为**细胞质多聚腺苷酸化元件**（cytoplasmic poly-adenylation element，CPE）。

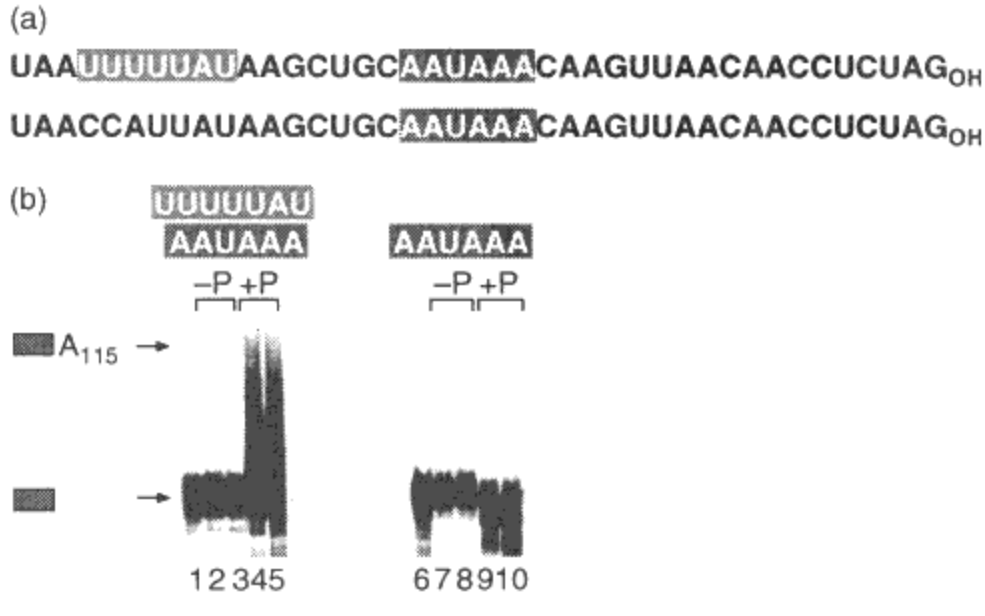


图 15.28　UUUUUAU 序列引发成熟特异性多聚腺苷酸化的证明。Wickens 及同事进行了图 15.27 的同样实验，利用同样的 SV mRNA，其 AAUAAA 基序的上游有或无 UUUUUAU 序列。(a) 两种待注入的 RNA 序列，其中 AAUAAA 和 UUUUUAU 被特别突出。(b) 结果。泳道 2～5 是卵母细胞 RNA，注入具有 AAUAAA 序列和 UUUUUAU 序列的 RNA。泳道 7～10 是卵母细胞 RNA，并注入只有一个 AAUAAA 序列。培养过程中是否具有孕酮在图 15.27 中已标记出。泳道 1 和 6 没有注入 RNA，左边的标准物与图 15.27 中的相同。UUUUUAU 基序是多聚腺苷酸化所必需的。（*Source*：Fox et al.，*Genes & Development* 3（1989）p. 2155，f. 5. Cold Spring Harbor Laboratory Press.）

AAUAAA 序列也是细胞质多聚腺苷酸化所需要的吗？为了回答这个问题，Wickens 及同事对 AAUAAA 序列进行点突变，然后将突变的 RNA 注入卵母细胞细胞质。他们发现，将 AAUAAA 序列突变成 AAUAUA 或 AAGAAA 序列，都会完全抑制多聚腺苷酸化。因此，这个基序是细胞核及细胞质多聚腺苷酸化都需要的。

小结 细胞质中非洲爪蟾母系 mRNA 的成熟特异性多聚腺苷酸化依赖于两个基序，接近 mRNA 末端的 AAUAAA 序列和称为细胞质多聚腺苷酸化元件的上游序列，后者是一个 UUUU-UAU 序列或与之类似的序列。

15.3 mRNA 加工事件的协同运作

了解 mRNA 的加帽、多聚腺苷酸化以及剪接之后，我们不难发现这些加工都是相互关联的。尤其是帽子为剪接所必需，但却仅对剪接第一个内含子是必需的。同样，poly（A）是剪接最后一个内含子所必需的。首先讨论一下加帽对剪接的作用。

帽子对剪接的作用

1987 年，Yoshiro Shimura 小组通过体外转录小鸡 δ-晶体基因片段获得了一系列标记的剪接底物。其中一个片段（图 15.29）含有被内含子隔开的外显子 13、14 和 15。为了给 RNA 加帽，他们在 m^7GpppG 或 GpppG 存在条件下转录了该基因片段。m^7GpppG 或 GpppG 引发体外转录，并在转录过程中添加甲基化或未甲基化的帽子。为了获得不加帽的 RNA，他们做了一个没有引物的体外转录。然后将标记的、加帽的和未加帽的剪接底物加入到 HeLa 细胞核提取物中，让剪接发生数次，最后，对混合物进行电泳，从提取物中分离产物和中间代谢物。

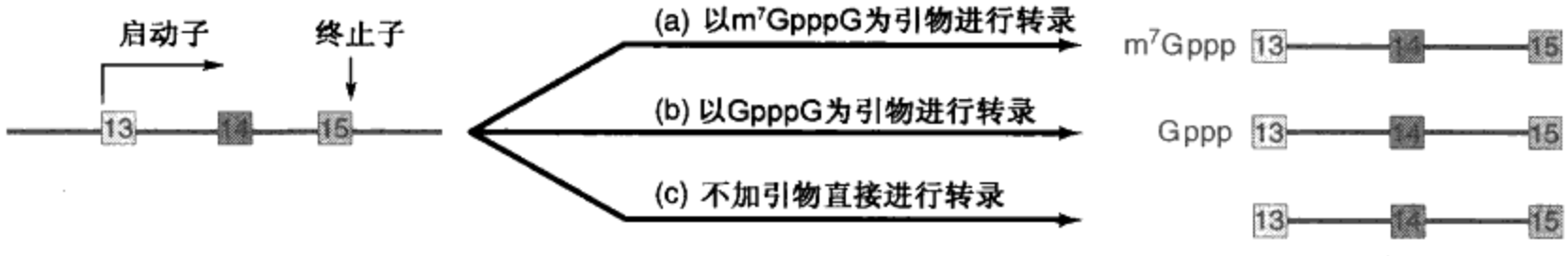


图 15.29 有帽子和无帽子剪接底物的获得。Shimura 及同事克隆了外显子 13、14 和 15，并连同其天然内含子插入载体，使其位于大肠杆菌启动子和终止子之间。他们在 3 种状态下对该基因进行转录，以获得 3 种不同剪接底物。(a) 以 m^7GpppG 引发并转录，以获得具有甲基化帽子的 RNA；(b) 以 GpppG 引发并转录，以获得具有未甲基化帽子的 RNA；(c) 不加引物转录以获得无帽子的 RNA。

图 15.30 介绍了实验结果。在加帽的前体中（图 15.30a），中间产物是两个内含子之一被剪接，但实验结束时其主要片段是完全剪接的产物。值得一提的是，含有未甲基化帽子的前体也能进行同样的剪接过程。但在未加帽的前体中，情况却完全不同。极少量的中间产物剪切了第一个内含子，大量的中间产物第二个内含子被剪切，未检测到完全剪接产物。可见，帽子的缺失似乎是选择性地抑制了第一个内含子的剪接。

未甲基化的帽子和甲基化的帽子在促进剪接作用上真的具有相同效果吗？可能并非如此，因为其他的实验表明，未甲基化帽子底物的剪接可被一种帽子甲基化抑制剂所抑制，即 S-腺苷高半胱氨酸（S-adenosyl homocysteine）。换句话说，帽子的甲基化似乎发生在剪接提取物中，尽管剪接提取物中并没有加入甲基化试剂。对帽子类似物抑制效果的研究进一步肯定了这一结论。研究发现，甲基化帽子对剪接的抑制作用比未甲基化帽子更有效。

帽子之所以能激发第一个内含子剪接，是因为第一个内含子最接近帽子，还是因为这个内含子有什么特殊序列使之依赖于帽子呢？为了寻找答案，Shimura 及同事用顺序颠倒的内含子合成了另一个剪接底物，即 14～15 外显子间的内含子换成第一个，而 13～14 外显子间的内含子换成第二个。帽子再一次剪接掉第一个内含子，而没有剪接掉第二个内含子。因此内含子和其邻接的外显子的顺序与剪接不相关，关键在于它接近帽子，帽子激发了与其最接近内含子的剪接。

1994 年，Mattaj 小组对帽子结合复合物（cap-binding complex，CBC）的特征进行了分析，此复合物有助于解释帽子对剪接的作用。CBC 由两个帽子结合蛋白（cap-binding protein，CBP）构成，其分子质量分别是 80kDa 和 20kDa，记作 CBP80 和 CBP20。抗 CBP80 的抗体可以免疫沉淀整个复合物（含有两种蛋白质），具有抗 CBP80 的 HeLa 细胞核提取物的免疫耗竭明显地抑制了含有一个内含子的模型底物的剪接。进一步实验证明，再次加入 CBC 后剪接活性又可恢复。相同的抗 CBP80 抗体也可以抑制剪接体前体（复合物 A、B、C 见图 14.27）的形成。因此，剪接体形成的第一步似乎是对帽子的

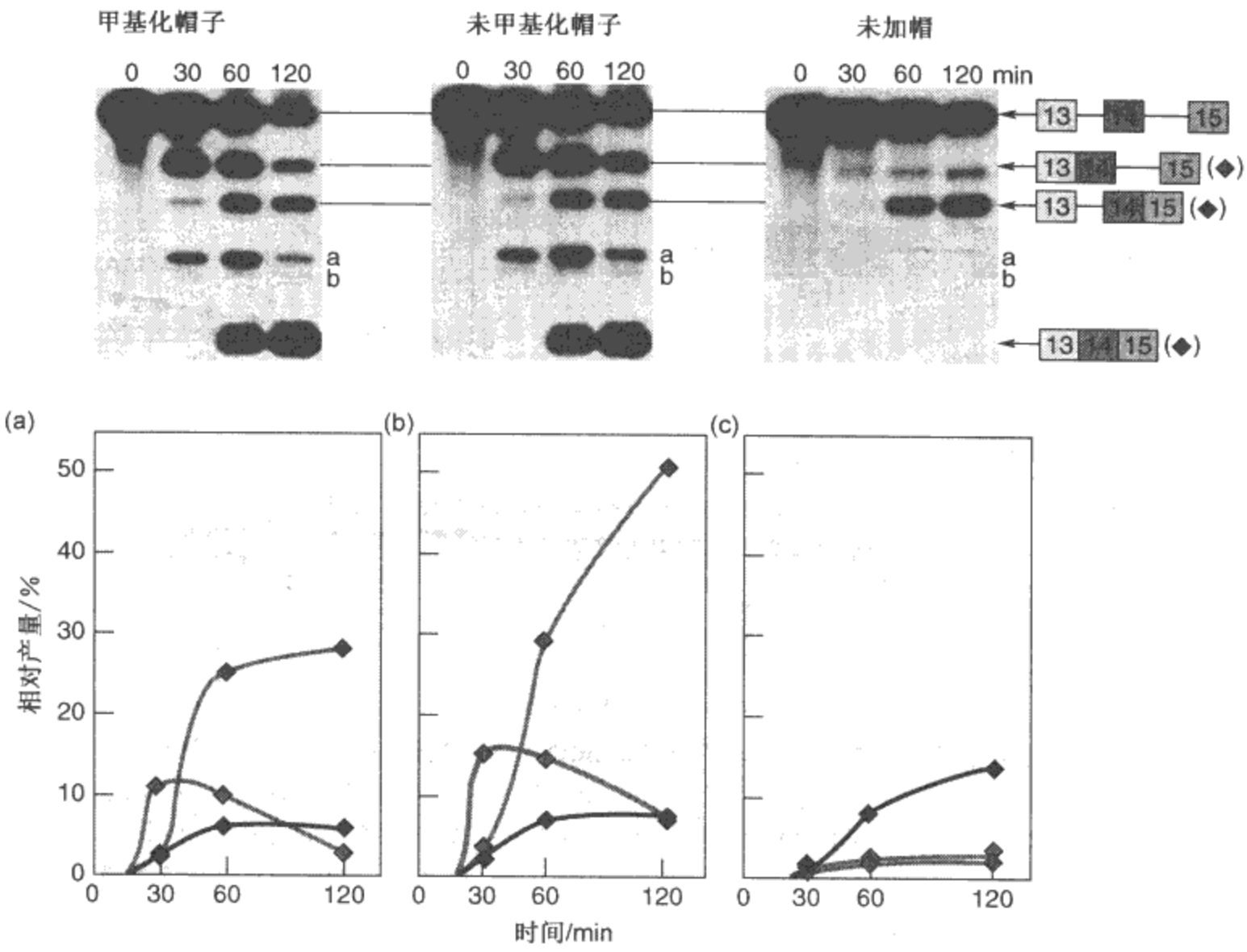


图 15.30　帽子在剪接含有两个内含子的底物时的作用。图 15.29 显示了 Shimura 及同事制作的^{32}P 标记的加帽和不加帽剪接底物模型，然后将这些底物与 HeLa 细胞核提取物孵育，时间如上图所示。电泳混合物，放射自显影检测剪接前体、中间物和产物。主要 RNA 的位置如图右侧所示，同时标出与下图相关的代表符号。条带 a 是第一个内含子的套索形式，条带是 b 含有外显子 13、14 的线形 RNA，第一个内含子在两个外显子之间。图（a）、（b）和（c）分别显示加帽甲基化、加帽、不加帽的前体底物的电泳结果。每个放射自显影图片下的曲线显示 RNA 的出现情况，图标与最右边的放射自显影标记相同。(*Source*: Adapted from Ohno, M., H. Sakamoto, and Y. Shimura, Preferentia excison of the 5′proximal intron from mRNA precursors with two introns as mediated by the cap structure. *Proceedings of the National Academy of Sciences USA* 84 (1987) p. 5189, f. 3.)

识别，而 CBC 对这一识别过程十分重要。

小结　在体外将第一个内含子从前体 mRNA 中切除依赖于帽子。这个作用可能参与剪接体形成的帽子结合复合物 CBC 所介导。

poly(A)对剪接的作用

有报道，多聚腺苷酸化在剪接前后都可发生，取决于所采用的系统。确实，有些前体 mRNA 的某些内含子在多聚腺苷酸化之前就剪接掉了，而有些内含子则在多聚腺苷酸化之后才被剪接掉。Maho Niwa 和 Susan Berget 用模式剪接底物证明了这个观点，并显示多聚腺苷酸化激活了前体 mRNA 中最后一个内含子而不是上游内含子的剪接。他们以 3 种标记的模型底物为对象，这些底物含有腺病毒主要晚期转录单元内含子（图 15.31a）。其中一个（WT）含有通常的多聚腺苷酸化信号（AAUAAA），因此可能被多聚腺苷酸化。第二个除了 AAUAAA 点突变成 AAGAAA 外 其余的和第一个完全相同，因此不会被多聚腺苷酸化。第三个（SB2）在其 3′端有一无关 RNA，无 AAUAAA 信号，因此不会被多聚腺苷酸化。将底物分别加入剪接提取物中，反应不同时间，然后电泳产物，放射自显影观察。图 15.31b 显示了实验结果。野生型底物有多种剪接，按测定的剪接表现分为：剪接的、未多聚腺苷酸化 RNA（S+）和剪接的、多聚腺苷酸化 RNA（A+S+）。图 15.31c 给出了图 15.31b 中剪接产物的量化结果。未多聚腺苷酸化底物的剪接速率是多聚腺苷酸化底物的 1/10～1/5。

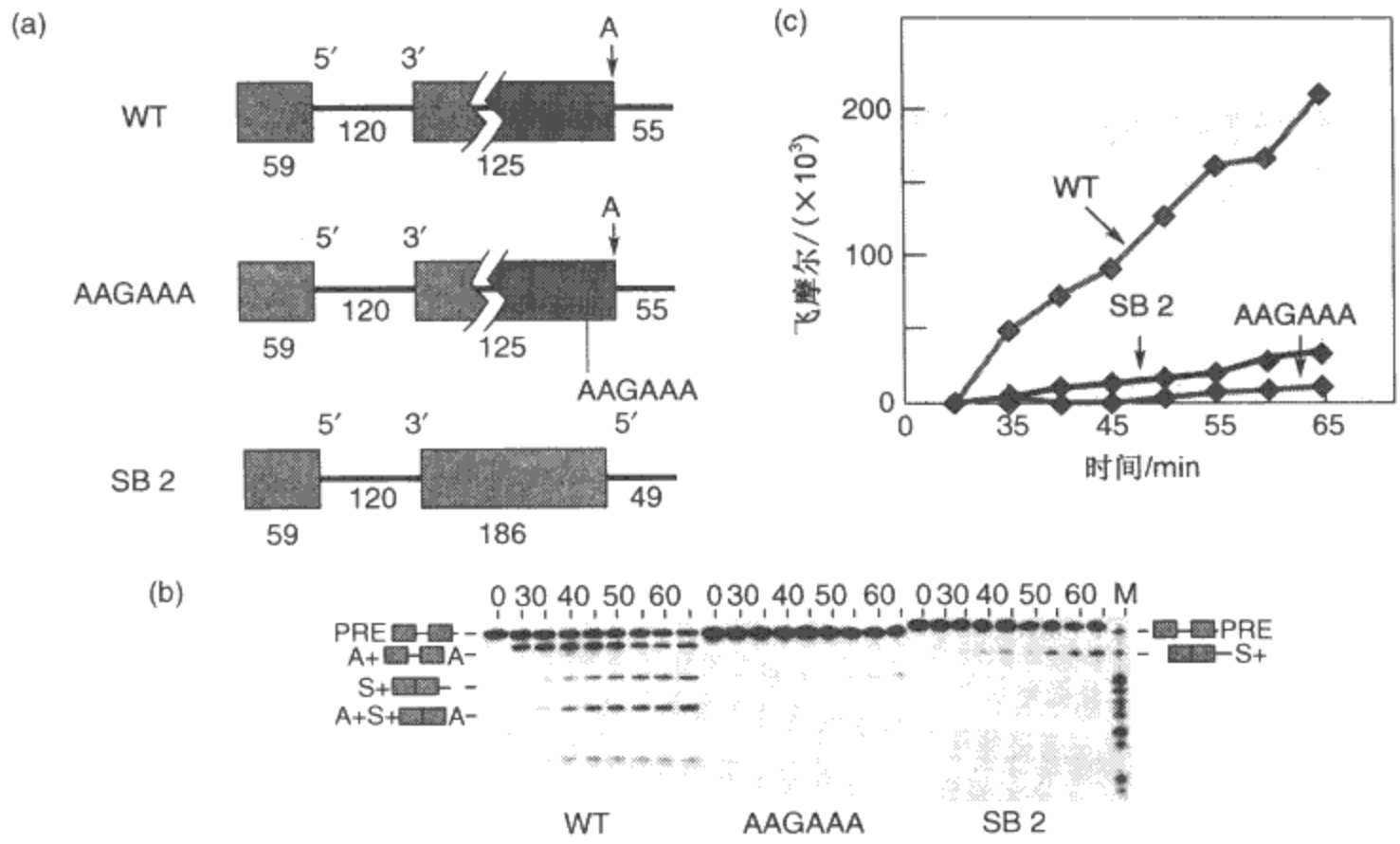


图 15.31　多聚腺苷酸化对只有一个内含子的前体 mRNA 剪接的影响。（a）底物设计。每个底物都有腺病毒主要晚期区的相同内含子。第一个（WT）含有野生型多聚腺苷酸化信号（AAUAAA）。第二个含有突变的多聚腺苷酸化信号（AAGAAA）。第三个（SB2）在其 3′端有不相关 RNA，但没有 AAUAAA 信号。(b) 剪接反应。Niwa 和 Berget 将标记底物在细胞核提取物中温育不同时间，然后电泳产物，放射自显影观察。底物在图底部标出，左右两边的示意图注明前体、中间产物、产物 RNA。Pre 代表前体 RNA，A＋是多聚腺苷酸化前体，S＋是剪接的未多聚腺苷酸化产物，A＋S＋是剪接的多聚腺苷酸化产物。(c) 剪接的量化。用作图仪确定 (b) 中每个剪接产物的条带中标记的量，并分析结果。绿色的是 WT RNA 的剪接，蓝色的是 SB2 RNA 的剪接，红色的是 AAGAAA RNA 的剪接。(*Source*: Niwa & Berget, Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns. *Genes & Development* 5 (1991) p. 2088. f. 1. Cold Spring Harbor Laboratory Press.)

之前我们了解到，帽子只影响其最邻近内含子（第一个内含子）的剪接。基于这个发现，我们也许期望 poly（A）只影响最靠近它的内含子（最后一个内含子）的剪接。为证明这一点，Niwa 和 Berget 制备了一对模式底物，带有两个相同的内含子，但加尾信号不同，一个是野生型 AAUAAA 信号（WT），另一个是突变的加尾信号（AAGAAA）。对这两个底物进行与前一个实验相同的处理。发现未多聚腺苷酸化底物的第一个内含子能够被很快地剪接，第二个内含子则不能。相反，多聚腺苷酸化底物的两个内含子都能很快被剪接掉。结果提示，去除末端的［靠近 poly（A）］内含子需要 poly（A），而去除远端［远离 poly（A）］内含子并不需要 poly（A）。

以含有 3 个完全相同（大小不同）内含子的 RNA 为底物的剪接实验，进一步证明多聚腺苷酸化只影响最后内含子的剪接。同样，缺失多聚腺苷酸化尾抑制最后一个内含子的剪接，但上游两个内含子不受影响。

小结　在体外，对最靠近 poly（A）的内含子实施有效剪接，需要模型底物多聚腺苷酸化。如果模型底物没有多聚腺苷酸化，仍能以正常速率从这些底物中剪接其他内含子。

Rpb1 的 CTD 与 mRNA 加工蛋白质的结合

在本章前面和第 14 章，我们已经了解到 mRNA 加工的 3 个事件——剪接、加帽和多聚腺苷酸化都发生在转录过程中。当新生 mRNA 只有 30nt 长时，加帽就开始了，此时 RNA 的 5′端才刚从聚合酶中露出。多聚腺苷酸化发生在 mRNA 多聚腺苷酸化位点被剪切之后，而此时的转录物仍然还在延伸。剪接则开始于转录的进行过程之中。我们也了解到加帽和多聚腺苷酸化至少分别激活了第一个和最后一个内含子的剪接。

参与所有这些加工过程的是 RNA 聚合酶 II 中

Rpb1 亚基的 CTD。在本章，我们已经知道 CTD 参与多聚腺苷酸化，但它在剪接和加帽中也发挥作用。事实上，已有直接证据表明，加帽、剪接和多聚腺苷酸化的酶都直接与 CTD 结合，CTD 为这 3 个加工事件提供了一个平台。

1997 年，David Bentley 及同事证明**加帽酶**（capping enzyme）与 CTD 相互作用。他们制作了含有 GST 的亲和柱，分别偶联野生型 CTD、野生型磷酸化 CTD、突变体 CTD、或者不连 CTD。然后让 HeLa 细胞提取物分别通过这些亲和柱检测柱结合洗脱物的鸟苷酸转移酶活性。将柱结合洗脱物与 [^{32}P] GTP 混合，[^{32}P] GMP 与鸟苷酸转移酶形成共价结合物（covalent adduct），将标记的酶通过 SDS-PAGE 和放射自显影检测。图 15.32 显示 GST 与 CTD 的结合情况，结果表明鸟苷酸转移酶只与磷酸化 CTD 结合。

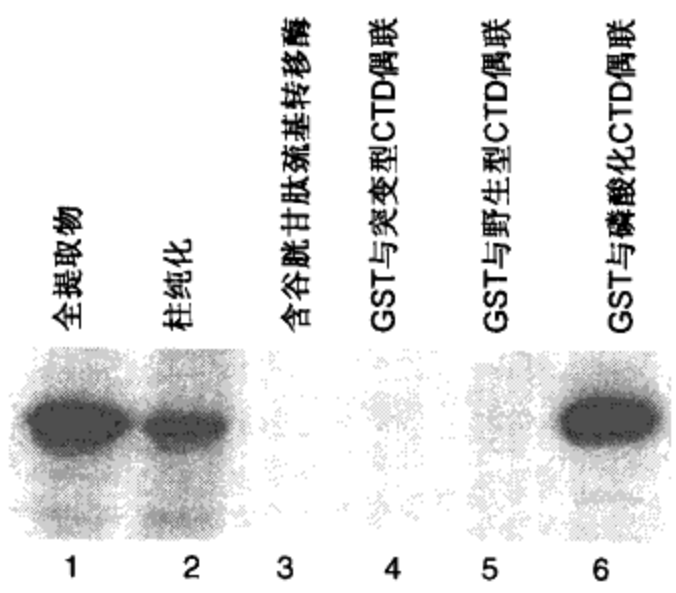


图 15.32　哺乳动物加帽谷胱甘肽巯基转移酶与磷酸化 CTD 的结合。Bentley 和其同事将 HeLa 细胞核提取物通过不同的亲和柱层析，纤维素基质所含底物如图上部所示，然后通过观察 [^{32}P] GMP 与酶共价结合物的形成来检测洗脱物中的谷胱甘肽巯基转移酶活性，结果以 SDS-PAGE 和放射自显影显示。L（泳道 1）为全提取物。FT（泳道 2）是柱的流出物。泳道 3～6 是流出物鸟苷酰转移酶实验结果。纤维素基质含 GST（泳道 3），GST 偶联突变 CTD（泳道 4），野生型 CTD（泳道 5），野生型磷酸化 CTD（泳道 6）。结果表明谷胱甘肽巯基转移酶只与磷酸化 CTD 结合。（*Source*：McCracken et al.，*Genes and Development* v. 11，p. 3310.）

2001 年，Nick 及同事用相似的方法证明了酵母剪切/多聚腺苷酸化因子 1A（CF1A）可结合磷酸化的 CTD。剪切/多聚腺苷酸化复合物的其他成分并不直接与 CTD 结合，而是与结合在 CTD 上的其他蛋白质紧密结合。此外，还有更多间接证据表明多聚腺苷酸化复合物与 CTD 的关系。研究发现，当 RNA 聚合酶缺失 CTD 时，不能很好地进行多聚腺苷酸化。CTD 尤其是其磷酸化形式，在体外激活多聚腺苷酸化。

参与前体 mRNA 剪接的蛋白质与 CTD 之间发生相互作用的实验证据也非常有力。Daniel Morris 和 Greenleaf 在 2000 年用**“Far Western 印迹”**实验证明酵母剪接因子 Prp40（U1 snRNP 的一个组分）与磷酸化 CTD 结合。Far Western 印迹实验与 Western 印迹实验相似，也是先将蛋白质进行 SDS-PAGE，然后印迹到硝酸纤维膜上。Western 印迹需用抗体进行标记，而 Far Western 印迹用可能与膜上某个蛋白质结合的另一个蛋白质进行标记。Prp40（也称为 WW 蛋白）电泳后印迹，用 [^{32}P] β-半乳糖苷酶-CTD 标记（CTD 与 β-半乳糖苷酶表达成融合蛋白，以便纯化。然后，体外磷酸化标记）。WW 蛋白具有两个色氨酸（W）结构域，频繁地参与 RNA 的合成与加工。

图 15.33 显示了上述分析的结果。图 15.33a 是考马斯亮蓝染色的胶，考马斯亮蓝是一种能与所有蛋白质结合的染料，因此图 15.33a 是一张全景的蛋白质图谱，显示胶上的所有蛋白质，包括 Prp40。每个泳道中分子质量最大的是目标蛋白，小的可能是其降解物。图 15.33b 是同一块胶的 Far Western 印迹结果，用 [^{32}P] β-半乳糖苷酶-CTD 标记。显然，Ess1、Prp40 和 Rsp5 都能与 CTD 结合。看来，仅有 WW 结构域不能保证其 CTD 结合活性，因为其他两个 WW 蛋白没有结合 CTD 探针。

小结　转录过程中，参与加帽、多聚腺苷酸化和剪接的蛋白质因子都与 CTD 结合。

RNA 加工蛋白与 CTD 结合的变化与 CTD 磷酸化相关

我们已经知道，三类主要 mRNA 加工蛋白都可与 CTD 结合，由此引出一个问题：已知 CTD 很长，可以同时与多种蛋白质结合，但它能同时跟参与 3 个加工事件的所有蛋白质和 RNA 都结合吗？

答案是根据即时任务需要，这些蛋白质可随时结合或离开 CTD。而且，蛋白质的结合或解离与转录时 CTD 磷酸化的改变相关。Steven 小组研究了加帽酶和多聚腺苷酸化酶与酵母多聚酶 II 的结合情况，他们分别检测了多聚酶 II 靠近启动子（起始后很短时间）和远离启动子（延伸期，起始后较长时间）时的结合情况，以及这两个时间 CTD 磷酸化的状态。

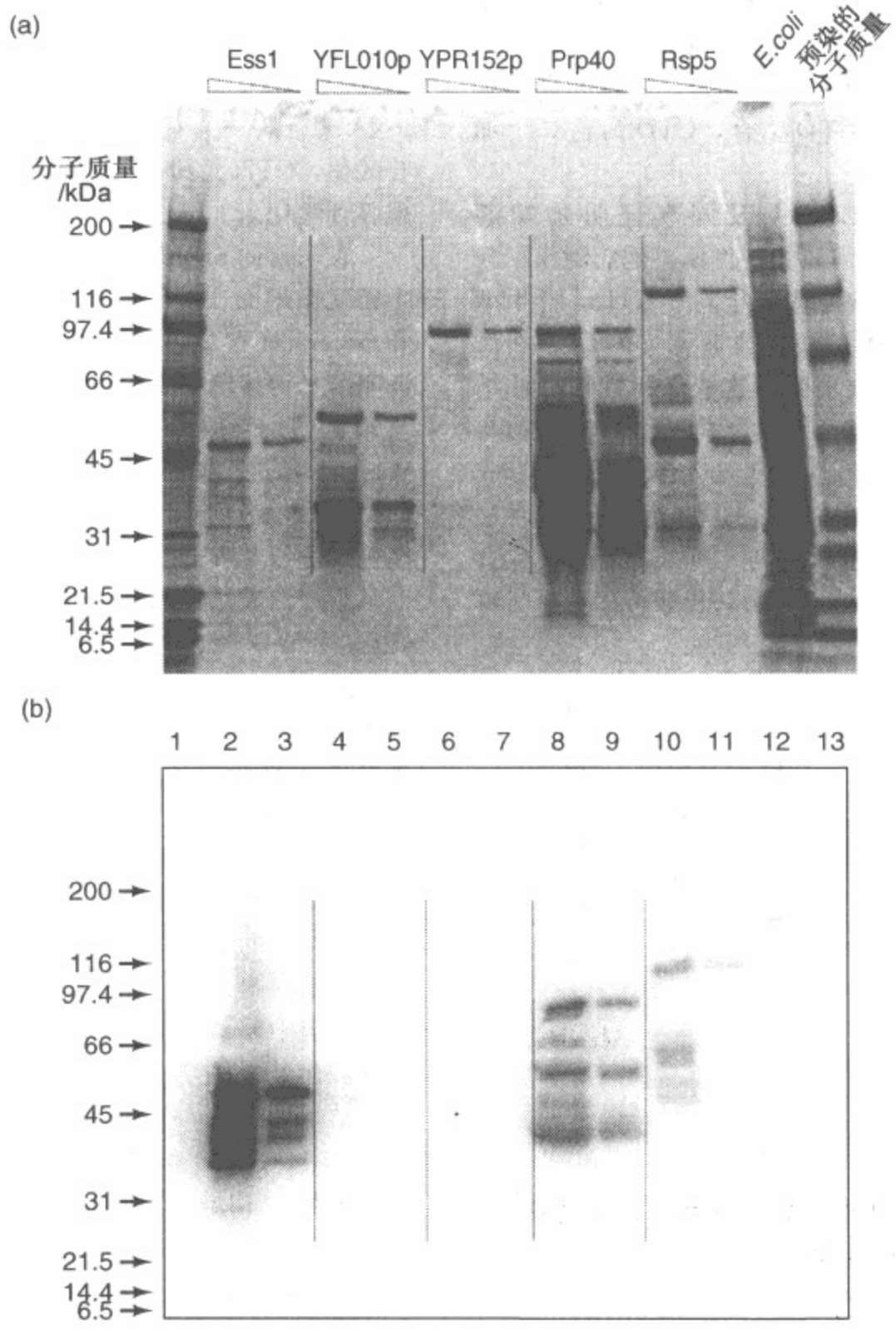


图 15.33　Prp40（以及其他蛋白）与 Rpb1 的 CTD 的相互作用。（a）凝胶电泳。Morris 和 Greenleaf 将已知具有 WW 结构域的 5 个蛋白质经 SDS-PAGE 及考马斯亮蓝染色。偶数泳道（2、4、6、8、10）含有 500ng 蛋白质，如图所示。奇数泳道（3、5、7、11）含有 50ng 相同蛋白质。每个泳道最上面的条带是完整目标蛋白。泳道 1 和 13 为标准蛋白，泳道 12 为大肠杆菌蛋白。(b) Far Western 印迹分析。图（a）的同一块胶被印迹到硝酸纤维素膜上，并用 [^{32}P] β-半乳糖苷酶-CTD 标记，磷屏成像。（*Source*：*Journal of Biological Chemistry* by Morris and Greenleaf. Copyright 2000 by Am. Soc. For Biochemistry & Molecular Biol. Reproduced with permission of Am. Soc. For Biochemistry & Molecular Biol. in the format Textbook via Copyright Clearance Center.）

他们发现，加帽酶（鸟苷酸转移酶）与靠近启动子的 CTD 结合，但与位于基因内部的 CTD 没有结合。相比之下，无论 CTD 与启动子的距离远近，甲基转移酶和多聚腺苷酸化因子 Hrp1/CFIB 都可以与之结合。因此，这些因子在转录的起始和延伸过程中都存在于转录复合物中。此外，他们还发现，CTD 七聚体（heptad）的 Ser5 在复合物靠近启动子时被磷酸化，但在延伸的后期却不被磷酸化；而其 Ser2 的磷酸化与 Ser5 刚好相反，延伸过程（远离启动子）被磷酸化，早期（这时聚合酶仍靠近启动子）不被磷酸化。

为此，Buratowski 小组应用了第 13 章介绍的染色质免疫沉淀（ChIP）技术，他们用抗加帽蛋白的抗体和抗多聚腺苷酸化蛋白的抗体，免疫沉淀染色

质，以便获得与这些蛋白质相互作用的聚合酶所转录的染色质。然后通过 PCR 鉴定这些免疫沉淀的染色质，PCR 所用引物是扩增几个不同基因的靠近启动子或远离启动子区域的专一性引物。

从这个实验中得出的其中一个可能的结果是：用抗特异性蛋白质的抗体所免疫沉淀的染色质与扩增启动子附近的引物可产生很强的 PCR 信号，而与扩增基因内部引物所产生的 PCR 信号比较弱。说明该蛋白质与转录复合物的结合发生在转录起始或起始不久，而不是在延伸阶段。

图 15.34 显示了利用下列抗体的 ChIP 实验的结果，分别是抗酵母加帽酶鸟苷酸转移酶（α-Ceg1）的抗体、抗酵母多聚腺苷酸化因子（α-Hrp1）的抗体、抗酵母 RNA 聚合酶 II 的 Rpb3 亚基（α-HA-Rpb3）的抗体。以每个抗体的免疫共沉淀染色质进行 PCR，PCR 引物是扩增 3 个酵母基因的启动子区和基因内部区域的专一性引物，这 3 个基因分别是乙醇脱氢酶（ADH1）、细胞质 H^+ ATPase（PMA1）、多种药物抗性因子（PDR5）。对这 3 个基因所得结果一致，证明：①只有在靠近启动子时，鸟苷酸转移酶（加帽酶）才与转录复合物结合；②多聚腺苷酸化因子与远离或靠近启动子的转录复合物都可以结合，而 RNA 聚合酶 Rpb3 亚基在远离和靠近启动子的转录复合物中都出现了，正如所预期的一样。

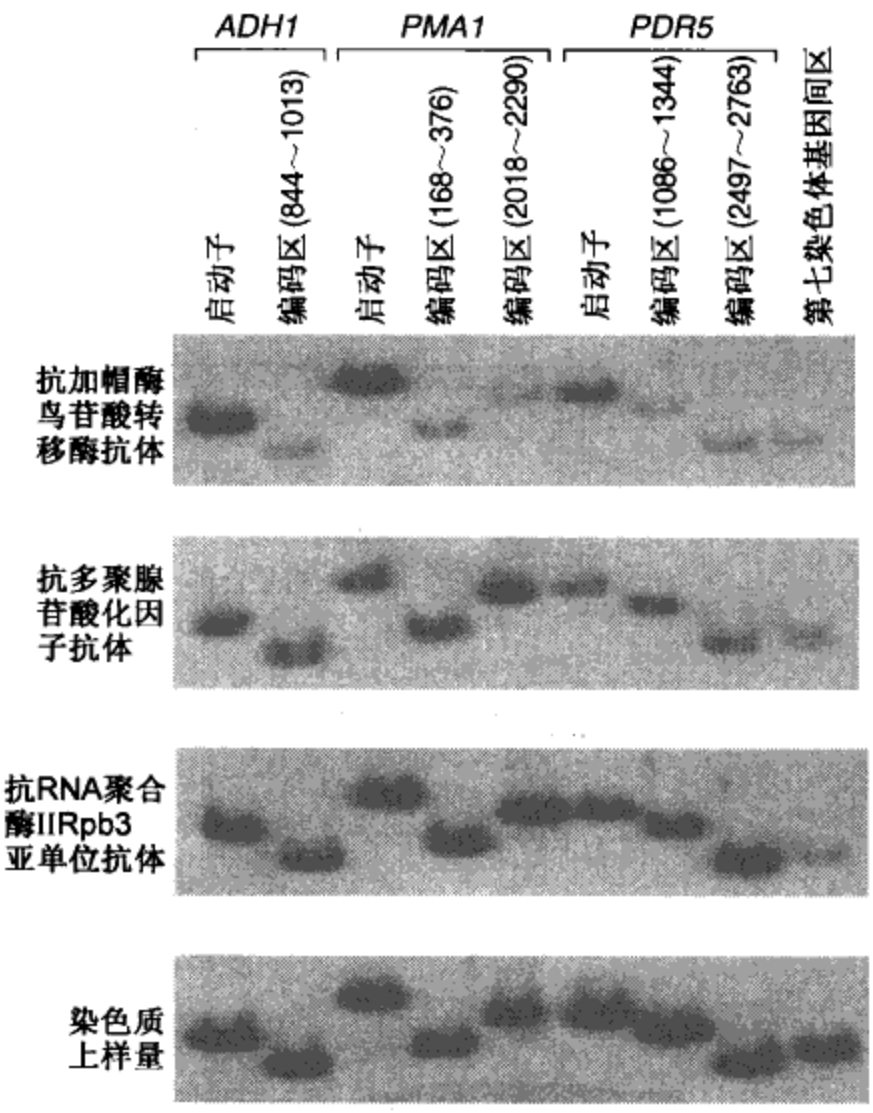


图 15.34　与 3 种酵母基因的转录复合物结合的蛋白质的 ChIP 分析。Buratowski 小组进行了 3 种蛋白质（加帽鸟苷酸转移酶、多聚腺苷酸化因子、RNA 聚合酶 II 的 Rpb3 亚基）与转录复合物靠近或远离 3 种不同基因（ADH1、PMA1、PDR5）的启动子时的结合情况的 ChIP 分析。用下列抗体免疫共沉淀染色质：抗加帽酶鸟苷酸转移酶（α-Ceg1）的抗体、抗多聚腺苷酸化因子（α-Hrp1）的抗体、抗 RNA 聚合酶 II Rpb3 亚基（α-HA-Rpb3）的抗体。用于每个实验的抗体在左边标出。对沉淀的染色质进行 PCR，PCR 引物可特异性地与 3 个基因的启动子区或内部编码区结合，以测定转录复合物是否靠近基因的启动子。信号强并有大量 PCR 产物表明靠近或远离启动子的相应的 DNA 存在于沉淀的染色质中。底部图是加入染色质的 PCR 结果，显示所有基因在免疫沉淀前均是等量加入。每个图的最后一个泳道是阴性对照，其引物可特异性地与染色体 VII 上、基因间未转录的区域结合，该区域存在于所加染色质中，但无抗体将之免疫沉淀。（*Source*：Reprinted by permission of S. Buratowski from "Komarnitsky, Cho, and Buratowski (2000) *Genes and Development* v. 14, pp. 2452-2460" © Cold Spring Harbor Laboratory Press.）

因此，通过 Rpb1 的 CTD 与转录复合物结合的蛋白质有一个动态变换。有一些只在转录过程的前期出现，有些出现的时间则更长些。什么原因使结合 CTD 的蛋白质谱发生改变呢？已经知道，转录过程中 CTD 的磷酸化状态会发生改变，因此可能是这一特性在起作用。

为了证明这种可能性，Buratowski 及同事用直接抗 CTD 七聚体重复中专一性磷酸化氨基酸（Ser2 和 Ser5）的抗体进行了 ChIP 实验。图 15.35 的 ChIP 实验发现，Ser5 的磷酸化主要发生在靠近启动子的转录复合物中，而 Ser2 的磷酸化发生于远离启动子的转录复合物中。随着转录复合物离开启动子，CTD 磷酸化很有可能从 Ser5 转换到 Ser2，导致了一些 RNA 加工蛋白（如加帽鸟苷酸转移酶）离开复合物，而吸引一些新蛋白质。图 15.36 对这个假说进行了总结。

小结　酵母转录复合物中 CTD Rpb1 的磷酸化状态随转录进程而改变。靠近启动子的转录复合物含有磷酸化的 Ser5，而远离启动子时，转录复合物含有磷酸化的 Ser2。与 CTD 结合的蛋白质也发生了变化。例如，当复合物靠近启动子时，加帽酶在转录早期而不是晚期出现；相反，多聚腺苷酸化因子 Hrp1 在靠近和远离启动子时，转录复合物都出现。

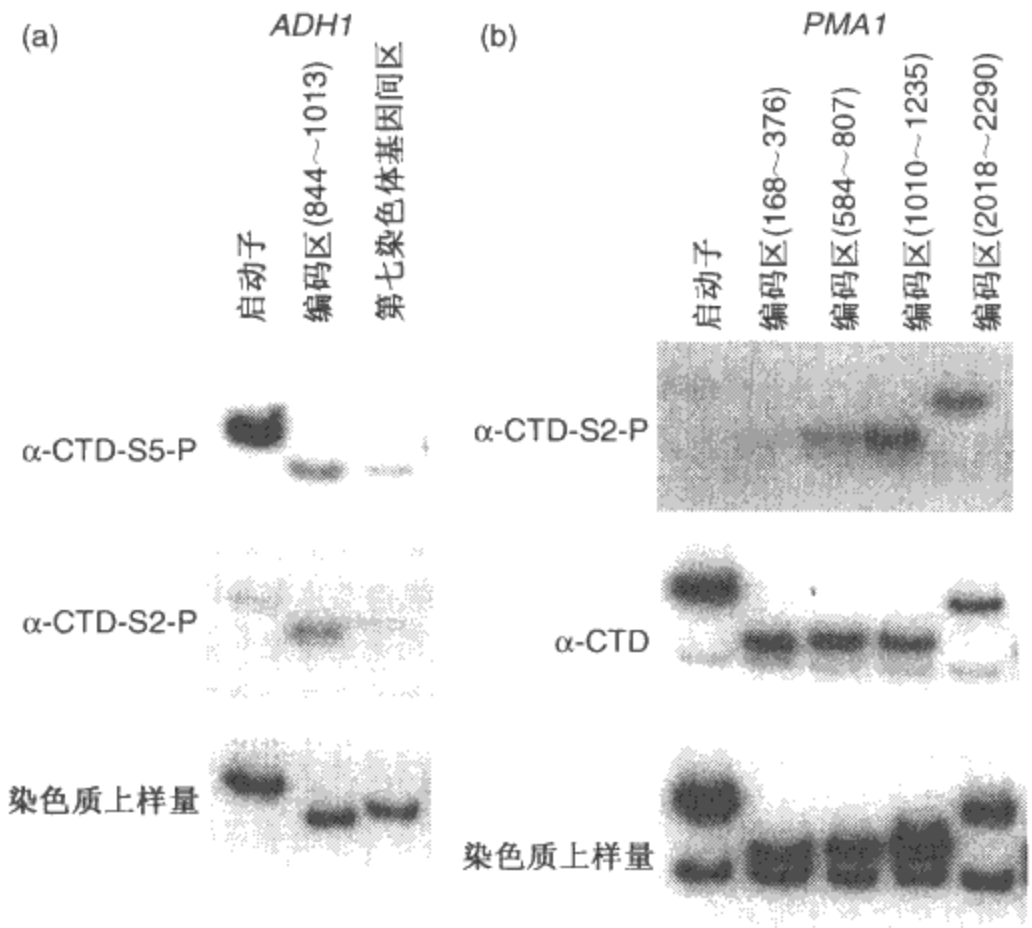


图 15.35　转录不同阶段 RNA 聚合酶 II 的 CTD 磷酸化状态的 ChIP 分析。Buratowski 及同事利用 ChIP 实验检测 RNA 聚合酶 II CTD Rpb1 亚基的两种磷酸化形式与染色质靠近或远离两个基因启动子的结合。(a) *ADH1* 基因的转录。用抗七聚体 Ser2 和 Ser5 都磷酸化的 CTD 的抗体免疫沉淀染色质，抗体在图的左边标出（分别为 α-CTD-S2-P 和 α-CTD-S5-P）。然后对沉淀的染色质进行 PCR，PCR 引物可特异性地扩增靠近启动子或远离启动子或基因间的区域，如图顶部所示。(b) *PMA1* 基因的转录。用抗 Ser2 上磷酸化或未磷酸化 CTD 的抗体免疫沉淀染色质，抗体在图的左边标出。PCR 引物如图顶部所示，可特异性地扩增启动子或逐渐远离启动子的区域（CDS 编码序列）。在两个图的底部都给出了染色质加样量对照。(*Source*: Reprinted by permission of S. Buratowski from "Komarnitsky, Cho, and Buratowski (2000) *Genes and Development* v. 14, pp. 2452-2460" © Cold Spring Harbor Laboratory Press.)

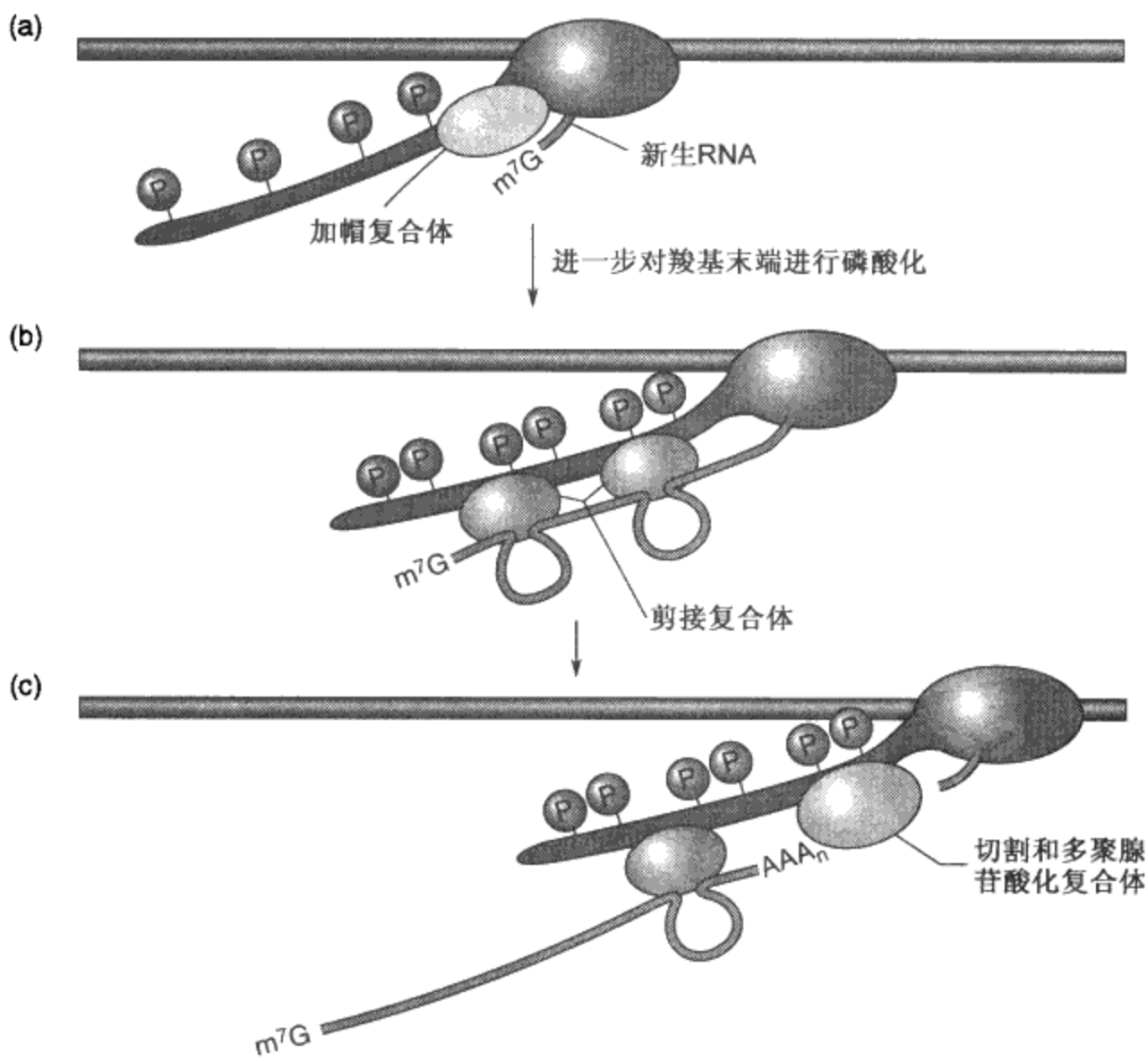


图 15.36　CTD 介导 RNA 加工过程的假说。(a) RNA 聚合酶（红色）已经开始合成新生 RNA（绿色）。部分磷酸化的 CTD 吸引了加帽复合物（黄色）。新 RNA 一转录出来，加帽复合物就给它加上一个帽子。(b) CTD 被进一步磷酸化（推测包括从 Ser5 磷酸化转换到 Ser2 磷酸化）并吸引了剪接复合物（蓝色），剪接复合物在转录时就界定外显子，并剪接外显子间的内含子。(c) CTD 结合剪切和多聚腺苷酸化复合物（橙色），该复合物可能在转录起始时就出现了，并开始剪切和多聚腺苷酸化。(*Source*: Adapted from Orphanides, G. and D. Reinberg, A unified theory of gene expression. *Cell* 108 [2000] p. 446, f. 3.)

转录终止与 mRNA 3′端加工的偶联

对 II 类基因的转录终止的研究十分困难，主要是因为成熟 mRNA 的 3′端与终止位点不一致。正如我们已经学过，长的前体 mRNA 必须在其多聚腺苷酸化位点被剪切，然后进行多聚腺苷酸化。这样就产生了一个相对稳定的 mRNA，但其 3′端却不稳定，会很快被降解。正是这个不稳定部分的 3′端是 mRNA 的终止位点。尽管很困难，但仍有几个研究者成功地研究了 II 类基因的转录终止，发现终止与多聚腺苷酸化位点的剪切相偶联，这两个过程相互依赖。确实，新生 RNA 在终止位点的剪切也许会早于在多聚腺苷酸化位点的剪切。

首先，我们如何知道转录终止与 mRNA 加工相偶联呢？Proudfoot 及同事通过酵母 II 类转录终止的研究证明了这种关联。特别是他们检测了酿酒酵母 *CYC1* 基因，发现参与剪切多聚腺苷酸化位点的蛋白质突变会抑制终止，而参与多聚腺苷酸化自身的蛋白质突变对终止没有影响。

Proudfoot 及同事将 *CYC1* 基因克隆到质粒（pCYC1）中，该质粒可以在强启动子 GAL1/10 下表达。同时还构建了一个相似的载体（pGcyc1-512），在其 *CYC1* 基因的末端缺乏通常的多聚腺苷酸化信号。然后用这些质粒转化酵母细胞，用 Northern 印迹检验基因的表达水平。图 15.37a 显示，多聚腺苷酸化位点的缺失极大地降低了基因的表达量。作为对照的看家基因（*ACT1*）在两个样品的表达量相同，表明 *CYC1* 信号的缺失不是由加样量不同或转膜不同造成的。

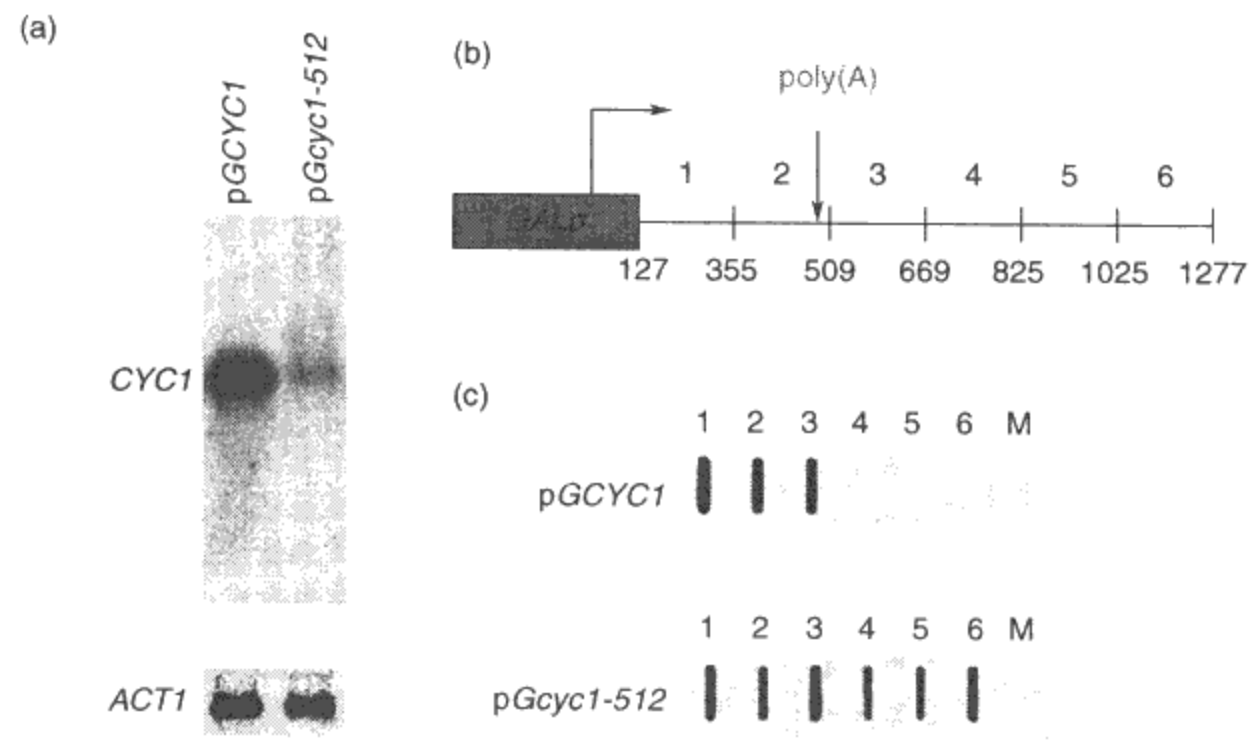


图 15.37　转录终止与多聚腺苷酸化的偶联。（a）Northern 印迹分析。Proudfoot 及同事将含有野生型基因（pCYC1）或缺失多聚腺苷酸化位点（pGcyc1-512）基因的细胞转录物进行 Northern 印迹。然后将印迹膜与标记的 *CYC1* 探针杂交。第一次杂交后，洗去印迹，用肌动蛋白基因探针（*ACT1*）再次杂交，作为印迹效率的对照。(b) 用于核连缀转录分析的区域图。将酵母 *CYC1* 基因克隆到质粒中置于强启动子 GAL1/10（GALp，绿色）的控制下，然后转化酵母细胞。进行核连缀分析时，斑点印迹长度和位置都已知的片段 1～6。片段 2 的多聚腺苷酸化位点已标出（红色）。(c) 连缀分析结果。用标记的转染了不同基因的细胞连缀转录物与片段 1～6（图 b）的斑点印迹膜杂交，转染的基因为野生型或突变 *CYC1* 基因，如图左边所示。M 是 M13 DNA，阴性对照。(*Source*：Birse et al.，*Science* 280：p. 299. © 1988 by the AAAS.)

pGcyc1-512 表达很弱的原因之一可能是没有合适地终止转录。为了确定终止是否真的失败了，Proudfoot 及同事进行了如下的核连缀分析（nuclear run-on analysis）。将 *CYC1* 基因的不同片段进行斑点印迹，其中包括多聚腺苷酸化位点下游约 800bp 的片段，如图 15.37b 所示。然后与转染了不同基因的细胞连缀 RNA 杂交，转染的基因为野生型 *CYC1* 基因或缺失多聚腺苷酸化位点的突变基因。图 15.37c 显示实验结果。野生型基因的转录终止于片段 3（多聚腺苷酸化位点的下游），因为没有与片段 4 杂交的转录物。但是突变基因的转录却远远超过了正常的终止位点，至少到了片段 6，表明正常终止失败了。

我们已经学过，多聚腺苷酸化实际上包含两个步骤：RNA 剪切和多聚腺苷酸化。理论上讲，两者之一可能与终止偶联。为了验证这个假设，Proudfoot 及同事进行了另一个连缀转录分析，所用酵母菌株编码剪切和多聚酰苷酸化因子的基因是温度敏

感突变型。同样，他们先做了 Northern 印迹，发现所有突变体在非允许温度下 *CYC1* mRNA 的表达量都很低。同样，转录物的多聚腺苷酸化失败或转录终止的失败都导致了转录物的不稳定性。

连缀转录分析给出了更为完整的答案。有些突变引起终止失败，而有些却不引起终止失败。其中是否存在某种模式呢？确实有！前一组基因（译者注：突变可引起终止失败的基因）编码的蛋白质参与多聚腺苷酸化之前的剪切，而后一组基因（译者注：突变不引起终止失败的基因）编码的蛋白质参与剪切后的多聚腺苷酸化。因此，看起来在多聚腺苷酸化位点的剪切而不是多聚腺苷酸化本身，与转录终止相偶联。

我们知道，剪切和多聚腺苷酸化因子与 RNA 聚合酶 II 中 Rpb1 亚基的 CTD 相关。转录终止过程需要活性剪切因子的结果提示，CTD 在终止以及 mRNA 的其他成熟过程中发挥作用。在下一节我们将回到这一话题。

小结 完整的多聚腺苷酸化位点以及在多聚腺苷酸化位点上进行剪切的活性因子都是转录终止所必需的（至少在酵母中是这样）。剪切过的 pre-mRNA 进行多聚腺苷酸化的活性因子并不是终止所必需的。

终止的机制

Michael Dye 和 Proudfoot 在 2001 年对人 β-和 ε-球蛋白基因的转录终止做了深入的研究。得到了以下结果：①多聚腺苷酸化位点的下游区域对于终止非常重要；②新生转录物中，在多聚腺苷酸化位点下游的不同位点进行剪切是终止所必需的；③这种转录物剪切是共转录性的（cotranscriptionally），推测可能先于在多聚腺苷酸化位点的剪切。后来，他们在 2004 年发现新生转录物的剪切是一个自我催化事件：RNA 分子可以自行剪切。

2001 年，Dye 和 Proudfoot 把含有 1.7kb 3′侧翼区的人 β-球蛋白基因转入质粒，该质粒由人免疫缺陷型病毒（HIV）的强增强子-启动子组合控制。然后用质粒转染可以表达人 β-球蛋白基因的 HeLa 细胞。HIV 增强子-启动子的优点是，其转录直接依赖于一个叫做 Tat 的病毒反式激活因子（transactivating factor），所以通过加入或移去 Tat 就能够很容易地使转录开启或关闭。

接下来，研究者对克隆的基因进行核连缀分析，并把实验结果与 β-球蛋白基因在天然染色体上由其自己的启动子控制所做的结果进行比较。图 15.38a 是 β-球蛋白基因，包含下游序列、自己的启动子及核连缀实验结果。从中可以发现，转录在经过位于多聚腺苷酸化位点下游 1.7kb 的 10 区后仍然能继续。图 15.38b 是克隆的受 HIV 增强子-启动子控制的 β-球蛋白基因的图谱和核连缀实验结果，从中看出转录在经过 10 区后仍然继续，但转录强度明显下降。10 区后的 DNA 包括载体的 A 区、B 区和 HIV 增强子-启动子的 U3 区。因此，终止发生在 10 区附近，转录和终止在构建的质粒中，似乎都能正常进行。

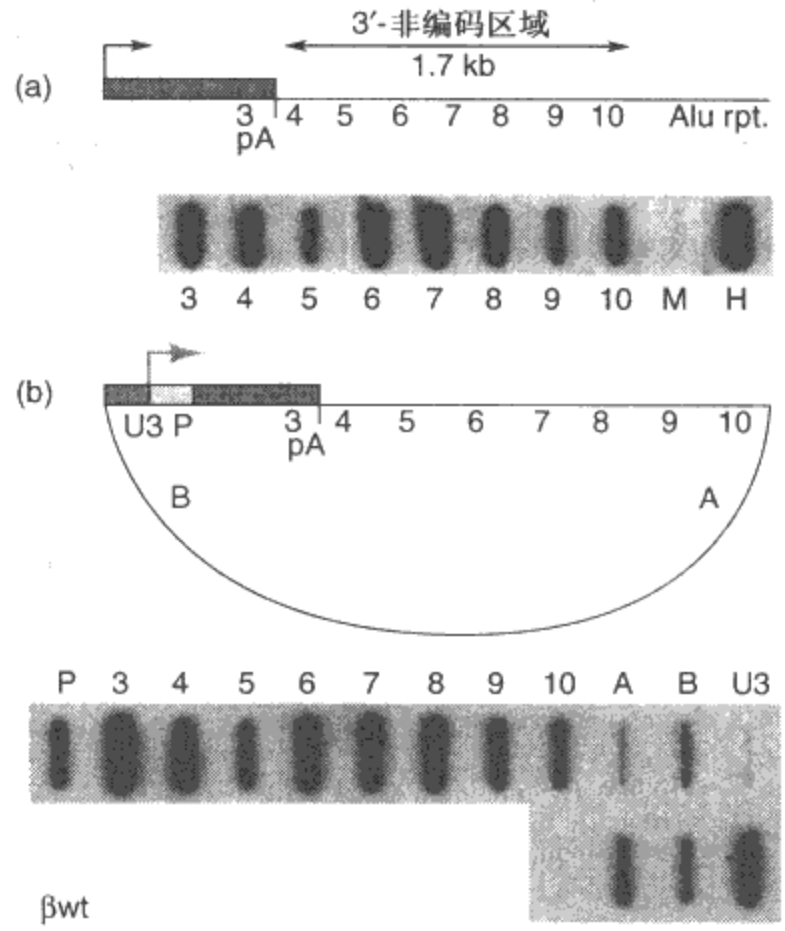


图 15.38 对于天然及克隆 β-球蛋白基因的核连缀分析。（a）基因在染色体上的情况。显示人的基因，包括启动子（紫色箭头所示为转录起点）、编码区（红色）、多聚腺苷酸位点（pA）及 1.7kb 的下游序列（4～10）区。下方的图是核连缀分析结果，包括 3～10 区及 2 个对照 M 及 H，其中 M 为阴性对照，是噬菌体 M13 DNA；H 为阳性对照，是人组蛋白 DNA，由 RNA 聚合酶 II 转录。(b) 在 HIV 增强子-启动子控制下的基因。HIV 增强子区（蓝色）、HIV 启动子区（黄色）、转录起始位点（紫色箭头）及编码区（红色），A、B 区属于克隆载体部分。该图下面是核连缀分析结果。M、H 的含义与图（a）相同，VA 代表腺病毒 VA1 基因与 β-球蛋白质粒共转染，由 RNA 聚合酶 III 转录。5S 表示与 5S rRNA 探针杂交，检测 5S rRNA 基因由 RNA 聚合酶 III 转录的情况。(*Source*: Reprinted from *Cell* v. 105, Dye and Proudfoot, p. 670 © 2001, with permission from Elsevier Science.)

之后，Dye 和 Proudfoot 缩小了 3′侧翼区里对转

录终止十分重要的区域。他们删除该区域的部分基因并用核连缀实验检测终止是否仍然发生。结果发现，一方面，删除 8～10 区可以阻止终止发生，所以 4～7 区不足以引起转录终止。另一方面，删除 5～8 区保留 9～10 区或者删除 5～9 区仅保留 10 区，转录也会终止。最值得注意的是，当删除 4 区下游的所有区域，只保留 8 区时，转录会终止。因此，区域 8、9 和 10，单独或在一起时都可以导致转录终止。

由于第 8 区（第 9、10 区也是如此）看上去含有能让合成中的转录物受到剪切的终止序列，Proudfoot 及同事将这段序列命名为**共转录剪切元件**（cotranscriptional cleavage element，CoTC 元件）。2004 年，Proudfoot 和 Alexander Akoulitchev 及同事发现了 CoTC 元件的重要秘密：它自身编码一个催化结构域，此结构域可剪切合成中的 RNA 分子。将含有全长 CoTC 元件的转录物与 Mg^{2+} 和 GTP 一起温育，发现该转录物的半衰期只有 38min，比对照 RNA 的降解快得多。通过在 CoTC 元件内部进行删除，他们将自身催化位点定位于 CoTC 元件 5′端下游的 200nt 序列［CoTC（r）］（图 15.39）。这段 200nt 序列的体外半衰期只有 15min。而含有 50～150nt 核苷酸的突变序列（*mut*Δ）没有自身催化活性。

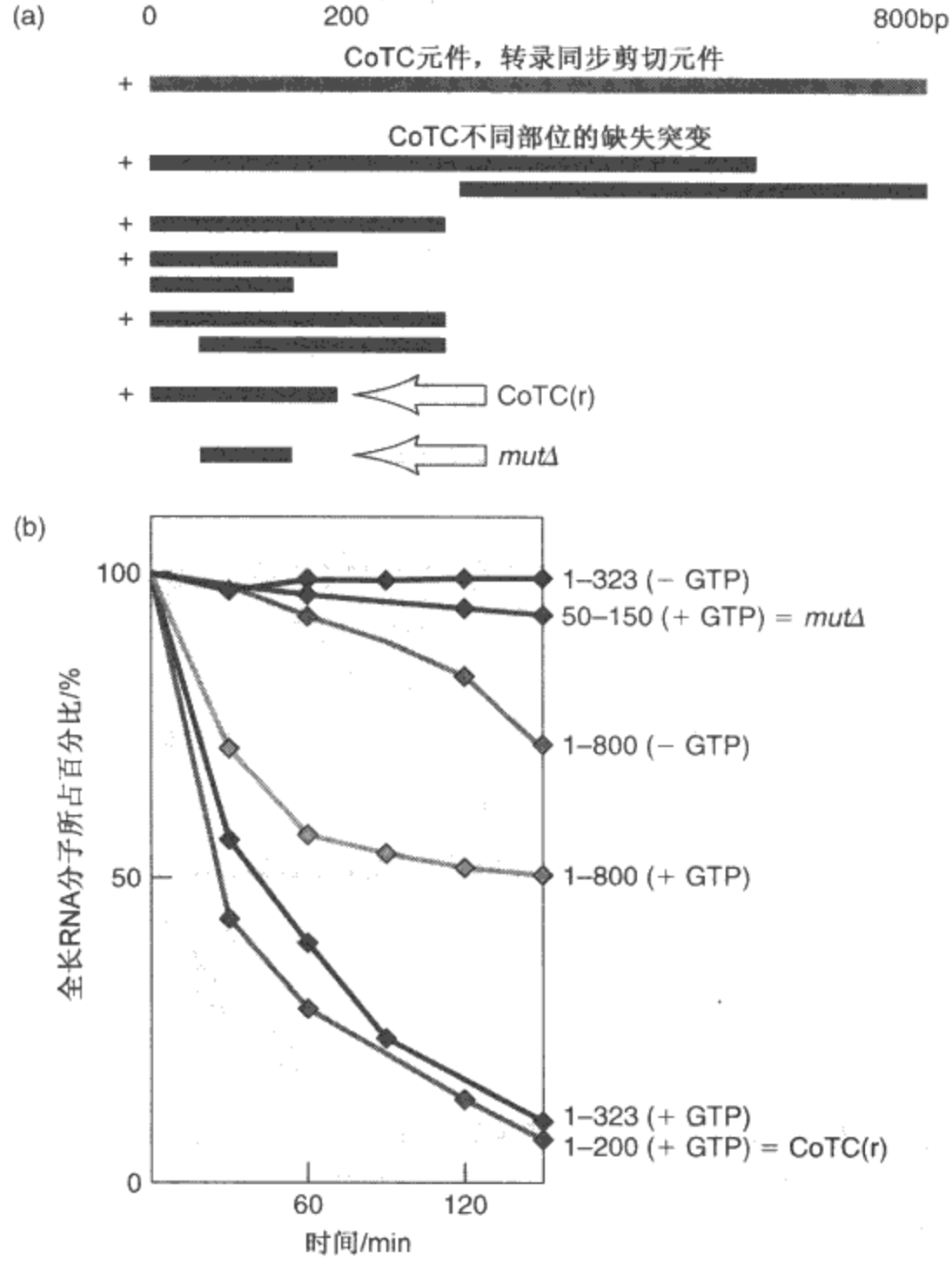


图 15.39　CoTC 元件内催化位点的确定。（a）突变体。Proudfoot 和 Akoulitchev 及同事从 800bp 的 CoTC 元件（顶部，红色柱）开始，进行删除突变，获得不同突变体，其转录物如下面（蓝色柱）所示。删除部分以柱中空缺表示。突变体 RNA 中仍保留催化活性在左边用＋表示。箭头指向 CoTC（r）：保留 1～200nt 的 RNA，具有催化活性。*mut*Δ：保留 50～150nt 的 RNA，不具催化活性。(b) 实验结果。将反应体系中所残留的全长 RNA 对反应时间作图。从图中可以看出反应过程需要 GTP，而含有 1～200nt 的 CoTC（r）具有完整催化活性。(*Source*：Adapted from A. Teixeira et al.，Autocatalytic RNA cleavage in the human beta-globin pre-mRNA promotes transcription termination. *Nature* 432：526，2006.)

CoTC 元件对转录终止很重要吗？为了找到答案，研究者将 β-球蛋白基因连接到质粒中，然后转染 HeLa 细胞。他们还用 β-球蛋白基因末端 CoTC 元件的突变体形式进行了替换，突变 CoTC 元件有 CoTC（r）（最小自身催化元件）和 *mut*Δ（缺失自身催化活性的元件）。然后用核连缀反应分析转录是否能正常终止。结果发现，末端含有 CoTC（r）元件的转录物能像野生型序列一样终止转录，而含有 *mut*Δ 元件的转录物在正常终止位点之后还能继续转录。用其他突变 CoTC 元件进行相同实验，发现 CoTC 元件的自身催化活性与其终止转录的活性非常吻合。由此看来，自身剪切活性对于正确的转录终止是必要的。

那么，是否真核基因的转录终止都需要一个能自身催化的类似 CoTC 的元件呢？研究表明，灵长类生物的 β-球蛋白基因确实含有一个保守的 CoTC 元件，其催化活性的核心区域内保守性最高。但是在亲缘关系稍远的物种没有检测到类似元件，可能是由于序列变异程度比较大的缘故。然而，仅从序列本身还不能将 CoTC 元件鉴定为自剪切性核酶（self-cleaving ribozyme）。因此，更多真核基因 poly（A）位点下游可能存在类似 CoTC 的元件。

延伸中的 RNA 分子在 CoTC 或其他位点的剪切就足以引起转录终止吗？可能不是，因为目前有证据表明另有一种现象，当 RNA 聚合酶延伸到 poly（A）位点之后，它会被“鱼雷”炸掉（torpedoed）。图 15.40 展示了该爆破过程的机制，与第 6 章学过的依赖 ρ 因子的转录终止机制很相似。首先，RNA 分子在 CoTC 或其他位点剪切，然后外切核酸酶结合到新形成的 RNA 末端开始降解，追赶仍在进行合成的 RNA 聚合酶。当外切核酸酶追上聚合酶时，对其进行“鱼雷式攻击”，终止转录过程。

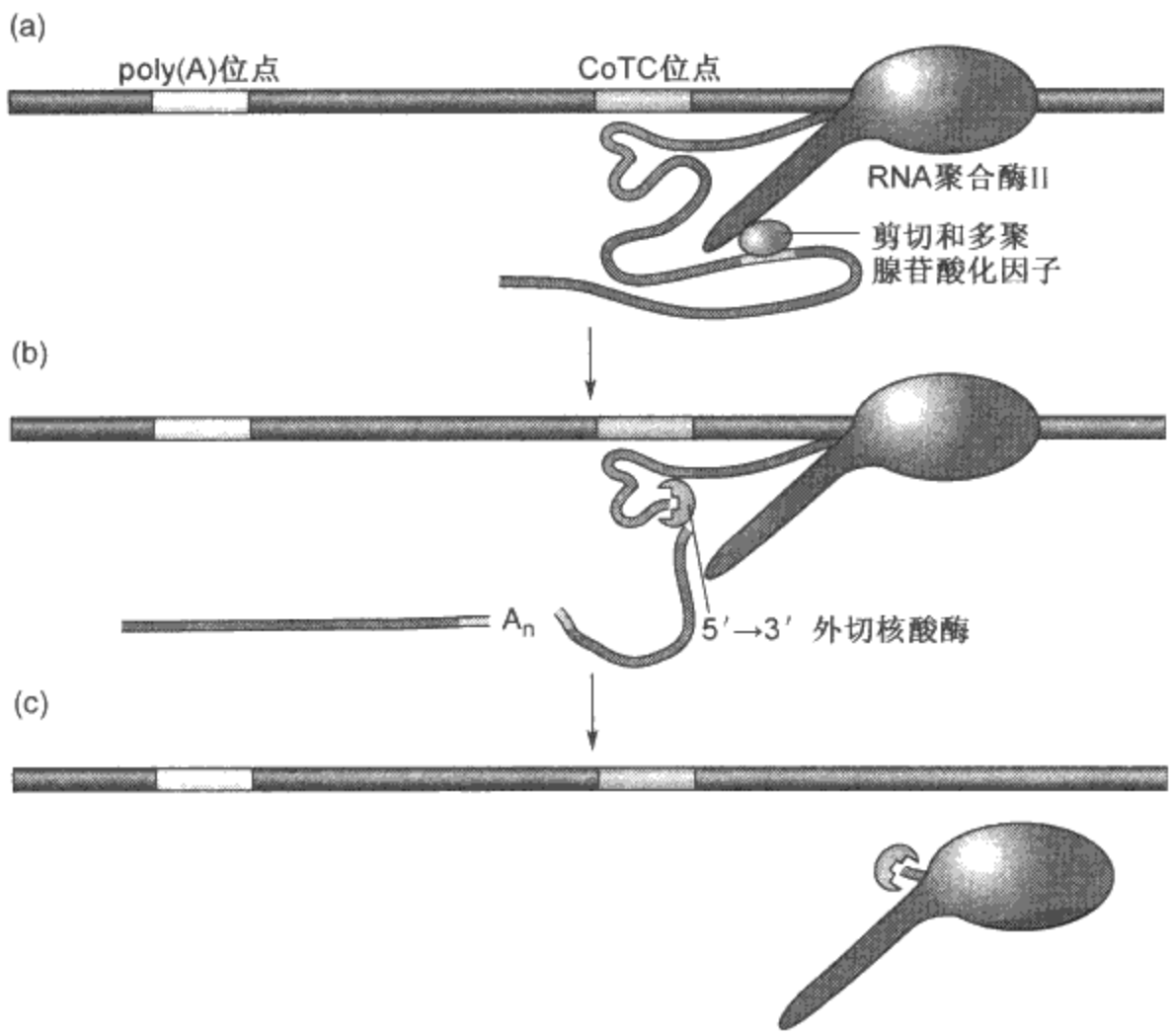


图 15.40　人 β-球蛋白基因转录终止的鱼雷模型。（a）RNA 聚合酶（红色）转录 poly（A）位点（黄色）和 CoTC 位点（蓝色）。剪切和多聚腺苷酸化因子（绿色）同时结合到 poly（A）位点和聚合酶的 CTD 上。（b）剪切和多腺苷酸化过程完成，mRNA 获得其 poly（A）尾巴。转录物的 CoTC 序列则经历自我剪切，而且 Xrn2 外切核酸酶（橙色）结合到新产生的 RNA 的 5′端。（c）Xrn2 一个核苷酸一个核苷酸地降解延伸中的 RNA，并追上 RNA 聚合酶，以某种方式将其破坏，使 RNA 聚合酶脱离模板，终止转录。

在人 β-球蛋白基因中，根据鱼雷式攻击模型，延伸中的转录物在 CoTC 位点剪切之后，为最终破坏聚合酶的 5′→3′外切核酸酶提供了一个登陆位点。如果确实如此的话，那么耗竭细胞中的相关 5′→3′外切核酸酶应该干扰转录的正确终止。为了验证这一点，Proudfoot 及同事利用 RNAi（第 16 章）对人

的5′→3′外切核酸酶 Xrn2 进行了表达“敲低”(knock-down)。他们将细胞内 Xrn2 的活性降低到只有正常值的25%，然后用核连缀实验对转录终止进行检测。发现 Xrn2 活性的降低使正常终止的转录物减少了二至三成。也就是说，有二至三成的转录物会延伸到正常终止位点之后。

Proudfoot 及同事认为在 poly（A）位点而不是在 CoTC 位点进行的剪切可能是 Xrn2 的登陆位点。如果真是这样，那么在 poly（A）位点和 CoTC 位点之间的 RNA 的耗竭在 Xrn2 抑制细胞内要低于未处理的细胞。但是 RNase 保护实验表明，在 Xrn2“敲低”细胞和未处理细胞内，从 poly（A）位点到 CoTC 位点之间转录物的水平并没有差异。

是否位于 CoTC 区域 5′端的任一位点都可以作为 Xrn2 的登陆位点呢？Proudfoot 及同事用榔头状核酶（hammerhead ribozyme）序列替代正常 CoTC 序列，希望能从中找到答案。这种榔头状核酶是可以自身剪切的 RNA，在末端生成 5′-OH，而 CoTC 自身剪切生成 5′-磷酸基团。核连缀分析表明，尽管榔头状核酶能够对延伸中的 β-球蛋白转录物进行剪切，但是其下游的 RNA 并没有被降解，正如细胞中含有正常 CoTC 序列的转录物一样。因此，为了启动对下游 RNA 序列的降解，Xrn2 需要由 CoTC 提供的 5′-磷酸基团。

对转录进行终止的鱼雷轰炸机制能广泛适用于其他生物吗？Jack Greenblatt、Steven Buratowski 及同事在酵母中发现了一个能够启动转录终止的5′→3′外切核酸酶，称为 Rat1。目前还没有证据表明酵母含有 CoTC 元件，因此只能假设转录物在 poly（A)位点被剪切之后，Rat1 就结合上去并开始追逐 RNA 聚合酶，追上之后对其进行爆破。

小结 RNA 聚合酶 II 引起的转录终止分两个阶段。首先，转录要在多聚腺苷酸化位点下游的终止区进行共转录剪切（CoTC）。这一步骤的发生早于 poly（A）位点的剪切及多聚腺苷酸化，并且不依赖于这一过程。第二步，在 poly（A）位点发生的剪切及多聚腺苷酸化给延伸中的 RNA 聚合酶发出脱离模板的信号。至少在某些基因内，这种脱离信号是由“鱼雷”传递的：对人的 β 球蛋白转录物而言，多聚腺苷酸化位点下游的 CoTC 元件是一种能够进行自我剪切的核酶，剪切之后产生一个 5′端。这种剪切对于转录终止是必需的，因为它为 Xrn2 提供了一个登陆位点。而 Xrn2 是一个 5′→3′的外切核酸酶，结合到 RNA 分子之后就开始一边降解 RNA 一边追赶 RNA 聚合酶。当它追上聚合酶后，对其进行“轰炸”，从而终止转录。酵母中也发现类似的鱼雷轰炸机制，poly（A）位点的剪切为 5′→3′的外切核酸酶 Rat1 提供了结合位点，Rat1 一边降解 RNA 一边追赶 RNA 聚合酶，追上之后就终止转录过程。

多聚腺苷酸尾在 mRNA 转运中的作用

自 1991 年以来，我们就知道多聚腺苷酸尾在成熟 mRNA 转运出核过程中发挥重要作用。当时 Max Birnstiel 及同事用实验证明当细菌的新霉素基因转至猴 COS1 细胞时，其转录物被留在核内。他们推测细菌基因缺少多聚腺苷酸信号，会形成无成熟 3′端的转录物，这可能是其不能转运到细胞质的原因。

为了证实这一假设，他们给新霉素基因加上了一个哺乳动物的 β-珠蛋白基因的强多聚腺苷酸化信号，可以使新霉素转录物多聚腺苷酸化。结果，转录物有效地从核转运到了细胞质。

2001 年，Patricia Hilleren 及同事对 poly（A）聚合酶基因温度敏感型的酵母突变株进行了研究。这些细胞在非允许温度下就停止新生转录物的多聚腺苷酸化。他们特别关注 *SSA4* 基因的转录物，这是个热激基因，当移至不允许温度时，其转录物就会积累。通过荧光原位杂交（FISH）显示，*SSA4* 转录物仍在核的小区域（small foci）内，靠近转录的部位。在野生型细胞或在允许温度下的突变细胞中，在核内检测不到这种转录物，推测它们经过多聚腺苷酸化后被转运至细胞质了。这些结果再次证实多聚腺苷酸化是 mRNA 主动转运出核所必需的，未加 poly（A）尾时，转录物甚至不会远离其转录位点。

小结 多聚腺苷酸化对于 mRNA 从核内的合成部位转运到细胞质是必需的。

总结

帽子是分步合成的。首先，RNA 焦磷酸酶将前体 mRNA 末端的磷酸基团去掉，之后鸟苷转移酶在此末端加上 GMP（来源于 GTP)。然后，两个甲基转移酶分别将鸟苷第 7 位的 N 和倒数第二个碱基 2′-OH 甲基化。这一系列事件发生在 RNA 链延伸到 30 个核苷酸之前的转录早期。对多数 mRNA 而言，帽子能够确保至少有些前体 mRNA 的正确剪接、促进成熟 mRNA 转运出核外、保护 mRNA 不被降解、增加 mRNA 的可译性。

绝大多数真核生物的 mRNA 及其前体在 3′端都有一段约 250nt 的多聚腺苷酸残基链。这段 poly（A）是在转录之后被 poly（A）聚合酶加上的。

poly（A）可以同时加强 mRNA 的稳定性和翻译能力，这两个方面的相对重要性因不同的系统而不同。

真核细胞基因的转录超出多聚腺苷酸化位点，然后转录物被剪切并产生被多聚腺苷酸化的 3′端。有效的哺乳动物多聚腺苷酸化信号的组成应该是在其前体 mRNA 的多聚腺苷酸化位点上游约 20nt 处有一个 AAUAAA 序列，接着 23～24bp 处有一个富 GU 区，紧接一个富 U 区。这些区域会有一定的变异，从而导致多聚腺苷酸化效率的变化。植物的多聚腺苷酸化信号通常也具有 AAUAAA 序列，只是这个序列在植物中比动物中有更高的变异性。酵母的多聚腺苷酸化信号更为不同，其中很少包含 AAUAAA 序列。

多聚腺苷酸化的发生需要前体 mRNA 的剪切和在剪切处的多聚腺苷酸化。在哺乳动物中，剪切反应的发生需要 CPSF、CstF、CFI、CFII、poly（A）聚合酶和 RNA 聚合酶 II（尤其是 Rpb1 亚基中的 CTD）。具有与新生 mRNA 类似 3′端的短 RNA 能够被多聚腺苷酸化。剪切底物的最佳多聚腺苷酸化起始信号是 AAUAAA 及其后连接的至少 8 个核苷酸。一旦 poly（A）达到 10nt，接下来的多聚腺苷酸化将不再依赖于信号 AAUAAA，而是依赖于 poly（A）本身。参与多聚腺苷酸化起始过程的两个蛋白质是 poly（A）聚合酶和结合到 AAUAAA 基序上的 CPSF。

poly（A）的延伸需要一个称为 poly（A）结合蛋白 II（PABII）的特殊因子。这个蛋白质结合到已经起始合成的寡聚 A 上，并通过 poly（A）聚合酶把 poly（A）延长到 250nt 或更长。PABII 独立于信号序列 AAUAAA 而起作用，它仅依赖于 poly（A），但其活性可被 CPSF 加强。

对小牛胸腺多聚腺苷酸酶可能是至少 3 种 cDNA 的混合物，由不同的选择性 RNA 剪接过程产生。根据这些序列预测的酶结构包括一个 RNA 结合区（RBD）、一个多聚酶模块（PM）、两个核定位信号（NLS1 和 2）和几个丝氨酸/苏氨酸富含区（S/T）。靠后的区域，但不是其余的，是体外酶活性非必需的。

poly（A）在细胞质中更新。RNA 酶使其缩短，poly（A）聚合酶再让它延长。当 poly（A）耗尽时，mRNA 也就面临降解了。细胞质中母系 mRNA 的成熟特异性多聚腺苷酸化依赖于两个基序——接近 mRNA 末端的 AAUAAA 序列和称为细胞质多聚腺苷酸化元件的上游序列（CPE），后者是一个 UUUUUAU 序列或与之类似的序列。

帽子和 poly（A）在剪接过程中发挥作用，至少在对最靠近 mRNA 两端内含子的切除中起作用。第一个内含子从前体 mRNA 中的切除依赖于帽子。这个作用可能被参与剪接体形成的帽子结合复合物 CBC 所介导。体外模型底物多聚腺苷酸化需要把最靠近 poly（A）的内含子有效地去除。但是从中去除其他内含子即使没有多聚腺苷酸化，仍能以正常速率发生。

转录过程中，参与加帽、多聚腺苷酸化和剪接的蛋白质因子都与 CTD 结合。酵母转录复合物 Rpb1 中 CTD 的磷酸化状态随转录进程而改变。靠近启动子的转录复合物含有磷酸化的 Ser5，而远离启动子的转录复合物含有磷酸化的 Ser2。与 CTD 结合的蛋白质也发生了变化。例如，当复合物靠近启动子时，加帽酶在转录早期而不是晚期出现；相反，多聚腺苷酸化因子 Hrp1 在靠近和远离启动子的转录复合物都出现。

完整的多聚腺苷酸化位点以及在多聚腺苷酸化位点上进行剪切的活性因子都是转录终止所必需的（至少在酵母中是这样）。剪切过的前体 mRNA 进行多聚腺苷酸化的活性因子并不是终止所必需的。RNA 聚合酶 II 引起的转录终止分两个阶段。首先，转录要在多聚腺苷酸化位点下游的终止区进行共转录剪切（CoTC）。这一步骤的发生早于 poly（A）位点的剪切及多聚腺苷酸化，并且不依赖于这一过程。第二步，在 poly（A）位点发生的剪切及多聚腺苷酸化给延伸中的 RNA 聚合酶发出脱离模板的信号。至少在某些基因内，这种脱离信号是由“鱼雷”所传递的：对人的 β-球蛋白转录物而言，多聚腺苷酸化位点下游的 CoTC 元件是一种能够进行自我剪切的核酶，剪切之后产生一个 5′端。这种剪切对于转录终止是必需的，因为它为 Xrn2 提供了一个登陆位点。而 Xrn2 是一个 5′→3′的外切核酸酶，结合到 RNA 分子之后就开始一边降解 RNA 一边追赶 RNA 聚合酶。当它追上聚合酶后，对其进行“轰炸”，从而终止转录。酵母中也发现类似的鱼雷轰炸机制，poly（A）位点的剪切为 5′→3′的外切核酸酶 Rat1 提供了结合位点，Rat1 一边降解 RNA 一边追赶 RNA 聚合酶，追上之后就终止转录过程。

复习题

1. 用 ^{3}H-AdoMet 和 ^{32}P 标记一个带帽的真核细胞 mRNA，然后用碱消化，接着将产物通过 DEAE 纤维素层析柱，图示与已知电荷寡核苷酸相比，帽子 1 的洗脱峰值。画出帽子 1 的结构并计算它所带的电荷。

2. 怎么才能知道帽子结构含有 7-甲基化鸟嘌呤核苷？

3. 列出加帽过程的步骤。

4. 描述加帽对 RNA 稳定性影响的实验并给出实验结果。

5. 描述加帽和多聚腺苷酸化对翻译协同影响的实验并给出实验结果。

6. 描述加帽对 mRNA 转运到细胞质的影响并说明实验结果。

7. 描述可以证明 poly（A）长度的实验并给出结果。

8. 如何知道 poly（A）是在 mRNA 的 3′端呢?

9. 如何知道 poly（A）是在转录后才加上的呢?

10. 描述实验并给出结果，显示 poly（A）对 mRNA 翻译效率、mRNA 稳定性、召集 mRNA 到多聚体上的影响。

11. 画出多聚腺苷酸化过程示意图，以延伸到多聚腺苷酸化位点之后的 RNA 为开始。

12. 描述实验并给出结果，证明转录并不是在多聚腺苷酸化位点停止。

13. 描述实验并给出结果，证明 AAUAAA 对多聚腺苷酸化的重要性。有哪些基序可以替代 AAUAAA 序列的作用? 相对于多聚腺苷酸化位点，这些基序是在什么位置找到的?

14. 描述实验并给出结果，证明 G/U 富集和 U 富集区对多聚腺苷酸化的重要性。相对多聚腺苷酸化位点而言，这些基序在什么位置?

15. 描述实验并给出结果，证明 Rpb1 的 CTD 对前体 mRNA 剪切的影响。

16. 描述实验并给出结果，证明 poly（A）和特异性因子 CPSF 的重要性。

17. 描述实验并给出结果，对于含有变异了的 AAUAAA 模体的底物而言，在其末端加上 40 个 A 对多聚腺苷酸化的影响。

18. 描述实验并给出结果，证明 CPSF 与 AAUAAA 结合而不是与 AAGAAA 结合。

19. 描述实验并给出结果，显示 CPSF 和 PABII 对带有 AAUAAA 或 AAGAAA 基序、有或无 poly（A）底物的多聚腺苷酸化的影响，你如何解释这些结果?

20. 绘出多聚腺苷酸化示意图，说明 CPSF、剪切酶、PAP、RNA 聚合酶 II 和 PABII 的作用。

21. poly（A）聚合酶 I（PAPI）的哪个部位是多聚腺苷酸化活性所必需的? 列出相关证据。

22. 描述实验并给出结果，鉴定细胞质多聚腺苷酸化元件（CPE）是细胞质多聚腺苷酸化所必需的。

23. 描述实验并给出结果，证明帽子在 pre-mRNA 的第一个内含子的剪接中发挥作用。

24. 描述实验并给出结果，证明 poly（A）影响 pre-mRNA 最后一个内含子的剪接。

25. 描述实验并给出结果，证明加帽酶连接到 RNA 聚合酶 II 的 CTD 处。

26. 描述 Far Western 印迹实验并给出结果，显示 U1 snRNP 的一个成分结合到 RNA 聚合酶 II 的 CTD 上。

27. 描述 ChIP 实验并给出结果，分析：（a）当加帽酶靠近而不是远离启动子时，它与 RNA 聚合酶 II 的 CTD 区结合；（b）当 RNA 聚合酶离开启动子时，CTD 的磷酸化状态发生改变。

28. 描述实验并给出结果，显示多聚腺苷酸化的失败会导致正确转录终止的失败。这一结果是由于多聚腺苷酸化本身导致的，还是由于不能在多聚腺苷酸化位点剪切而引起的?

29. 描述实验并给出结果，证明转录终止需要转录物的自行剪切，即使它仍在延伸（共转录剪接）。

30. 描述真核细胞终止转录的鱼雷爆破模型。

分析题

1. 假设你正在研究一种病毒，它所产生的 mRNA 分子含有一种特殊的帽子结构，其净电荷为－4 而不是－5。然后你发现这些帽子结构像 1 型帽子一样进行了甲基化，即 m^7G 和倒数第二个核酸 2′-OH，但其他位点没有被甲基化。请提出一个假说来解释负电荷减少的原因并设计一个实验来证明你的假设。

2. 请设计一个实验来证明 CstF 是与剪切和多聚腺苷酸化信号中的 GU/U 元件结合的。你如何确定 CstF 的结合是需要富 GU 区或富 U 区其中一个序列元件还是两者都不可缺少?

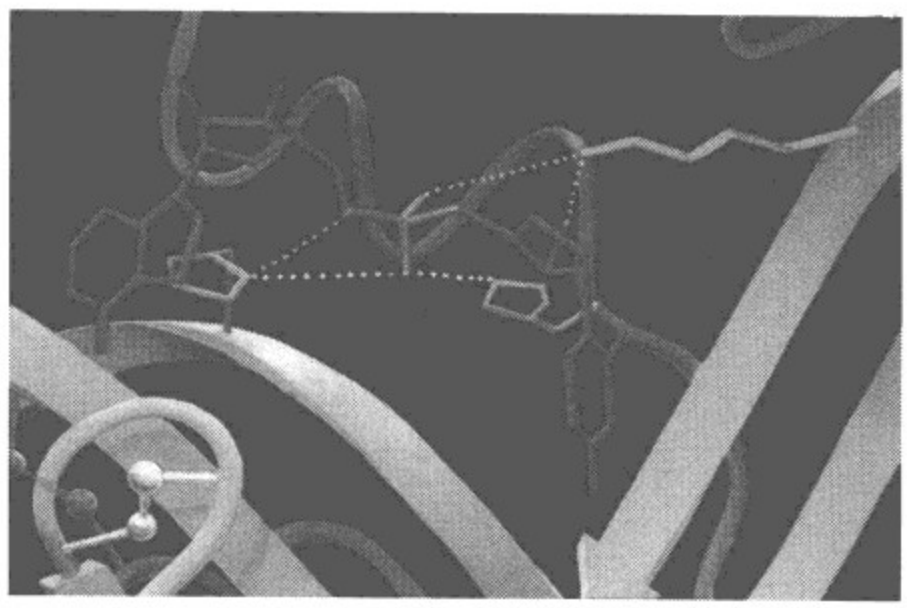
3. 假设你正在某一实验室从事 mRNA 加工的生化研究，通过体外实验对 mRNA 分子的剪切和多聚腺苷酸化过程进行分析。你合成了如下带放射性的 mRNA 分子（见下表）：含有 5′帽子或 5′帽子缺失。然后你将合成的 mRNA 分子与 HeLa 细胞核抽提物在 30℃中温育 20min，随后在高分辨率的凝胶上进行电泳。你根据分子质量的大小对剪切产物进行鉴别，然后对不同分子的放射量进行计数，包括没有加工的前体 mRNA、只有内含子 1 被剪切的分子（剪切 1）、只有内含子 2 被剪切的分子（剪切 2）、二者都被剪切的分子，还有被多聚腺苷酸化的分子［poly（A）］。结果如下表所示，其中加号的个数表示电泳条带放射性的相对强弱：

	pre-mRNA	剪切 1	剪切 2	二者都被剪切	poly（A）
未加帽的 RNA A	++	+	+++	+	+++
已加帽的 RNA A	+	+	+	+++	+++
未加帽的 RNA B	++++	+	+	+	+
已加帽的 RNA B	++	+++	+	+	+

请提出一个假说解释上面的实验结果。

翻译　肖海林　校译　刘智捷　岳　兵

第16章　其他RNA的加工过程



核糖核酸酶的计算机模型© *Irving Geis/Peter Arnold, Inc.*

在前两章，我们探讨了 RNA 的剪接、加帽和多聚腺苷酸化，它们涵盖了大部分真核细胞前体 mRNA（pre-mRNA）的加工过程。但在一些生物中还存在其他特殊的前体 mRNA 加工形式。例如，营寄生的原生动物锥体虫和蠕虫以及自主生活的原生生物眼虫等对前体 mRNA 进行反式剪接加工。这一过程涉及对两个独立转录物同时进行剪接。锥体虫含有线粒体，称作动基体，它们通过转录后增加或者删除核苷酸来编辑 mRNA。与这些相当复杂的加工过程相比，大多数生物则通过更加传统的机制加工 rRNA 和 tRNA。同时真核生物也可以通过调节基因转录后的加工，调控一些基因的表达，主要是通过调节 mRNA 的降解。此外，真核生物还可以通过降解相应的 mRNA 对外源基因或外源双链 RNA 做出反应。本章主要讨论这些转录后的修饰事件。

16.1　核糖体 RNA（rRNA）的加工

真核生物和细菌的 rRNA 基因首先被转录成较大的 rRNA 的前体，然后必须经过加工（切割成片段）才能形成成熟的 rRNA。然而，这个过程不仅是简单地移除过长的大分子两端所不需要的物质，那些埋藏于长的前体中的几种不同 rRNA 分子也都必须被剪除。下面分别讨论真核生物和细菌 rRNA 的加工过程。

真核生物 rRNA 的加工

真核生物的 rRNA 基因在细胞核中重复了几百次并成簇聚集在一起。已经对两栖动物中 rRNA 基因的排列进行了详细的研究，如图 16.1a 所示，它们被**非转录间隔区**（nontranscribed spacer，**NTS**）分开。NTS 与**转录间隔区**（transcribed spacer）不同，转录间隔区可作为前体 rRNA 的一部分被转录，然后在前体加工为成熟 rRNA 的过程中被除去。

细胞核内 rRNA 基因重复成簇的特点使得它们很容易被找到，这给 Oscar Miller 及同事提供了观察 rRNA 基因动态变化的机会。研究人员用电子显微镜观察两栖动物的细胞核，看到了惊人的现象，如图 16.1b 所示。从图中可以看到含有 rRNA 基因的 DNA 缠绕在一起，在显微照片中最显著的特点是一系列的“树”状结构。我们很快就要了解到这些转录物实际上是前体 rRNA，而不是成熟 rRNA 分子。“树”之间的间隔是非转录间隔区。你甚至可以根据一个基因中转录物的长度说出其转录方向，较短的 RNA 位于基因的起始端，而较长的位于基因的末端。

mRNA 前体常常需要剪接但不是其他的修剪（trimming）。而 rRNA 和 tRNA 首先以前体形式出现，该前体有时需要剪接，但它们的末端也含有多余的核苷酸，甚至在将成为单个成熟 RNA 序列的区域之间也存在多余的核苷酸。这些多余区域也必须切除。从前体 RNA 中修剪多余区域是另一种加工（processing）方式。它与剪接的相似之处是去除

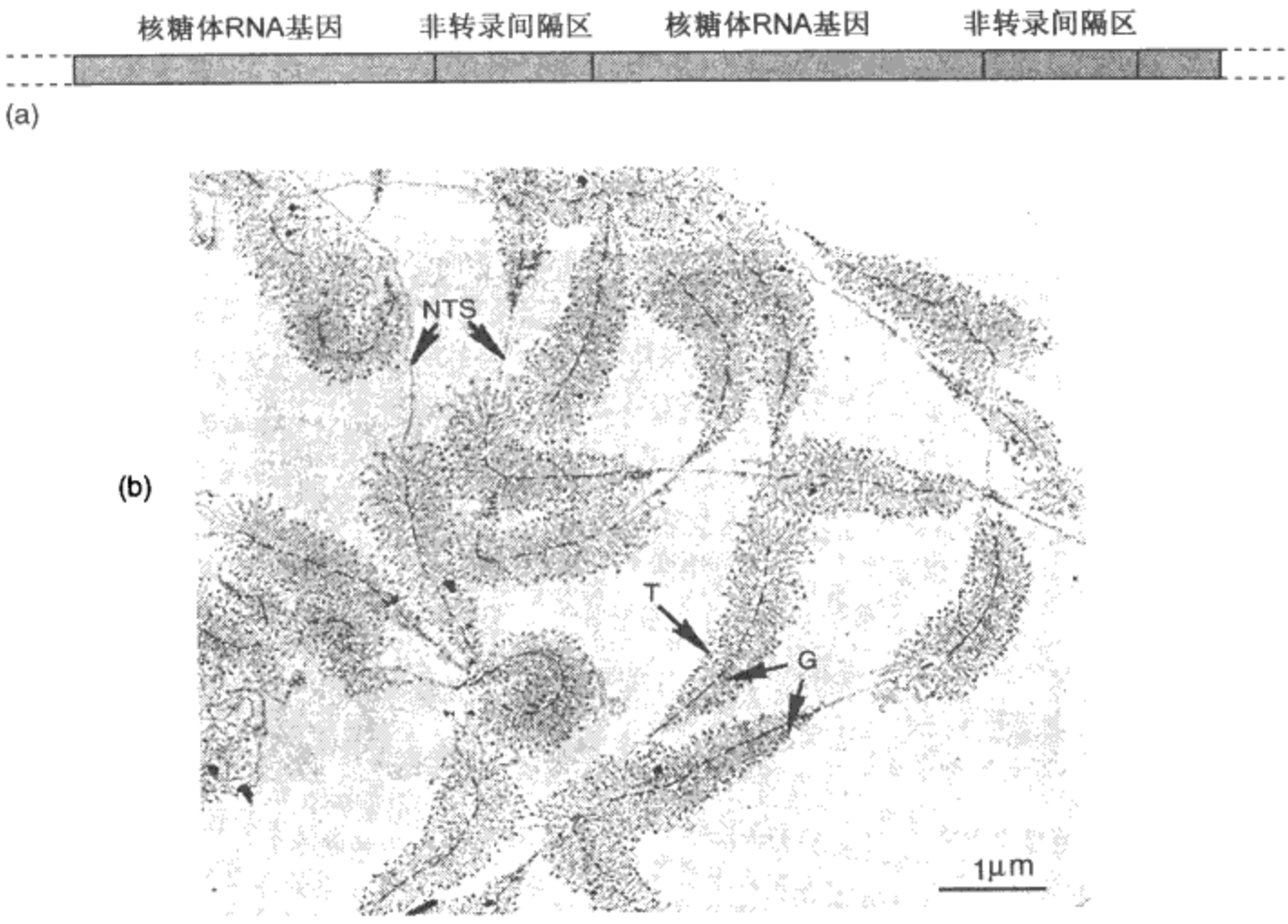


图 16.1　前体 rRNA 基因的转录。（a）蝾螈（两栖类）rRNA 前体基因簇的部分图谱。显示 rRNA 基因（橙色）和非转录间隔区（NTS，绿色）交替排列。(b) 蝾螈细胞核的部分电子显微图，显示前体 rRNA 转录物（T）以圣诞树的方式在串联重复的前体 rRNA 基因（G）处合成。在每个转录物的底部是 RNA 聚合酶 I（在本图看不到）。这些基因被非转录间隔区 DNA（NTS）间隔开来。(*Source*：（b）O. L. Miller, Jr., B. R. Beatty, B. A. Hamkalo, and C. A. Thomas, Electron microscopic visualization of transcription. *Cold Spring Harbor. Symposia on Quantitative Biology* 35（1970）p. 506.）

不需要的 RNA，但与剪接的不同之处是没有 RNA 连接。

例如，哺乳动物 RNA 聚合酶 I 合成一个 **45S rRNA 前体**，它包括 28S、18S 和 5.8S rRNA，它们被包埋在转录间隔区之间的 RNA 区域。该前体的加工过程发生在 rRNA 合成和核糖体聚集的细胞核中（图 16.2）。首先是切除 5′端的间隔区，留下一个 41S RNA 的中间体。接下来，41S RNA 被切割形成 32S 和 20S 两个片段，它们分别包含 28S 和 18S 序列，32S 中间体同时也含有 5.8S 序列。最后在 32S 中间发生剪切，产生成熟的 28S 和 5.8S RNA，它们之间可以碱基互补配对，同时 20S 中间体被修剪为成熟的 18S rRNA。

这些加工事件依次发生的证据是什么呢？早在 1964 年，Robert Perry 就用**脉冲示踪**（pulse-chase）建立了 45S 前体与 18S 和 28S 成熟 rRNA 之间的前体-产物关系。他用 [^{3}H] 尿苷酸短时间标记鼠 L 细胞（短脉冲），发现标记的 RNA 以 45S 为中心聚集成宽峰。然后“追踪”该 RNA 的标记进入 18S 和 28S rRNA。具体做法是，加入过量的非标记尿苷酸来稀释标记的核苷酸，然后观察到随着被标记的 45S 前体数量的减少，成熟 18S 和 28S rRNA 中标记的数量也减少。这说明 45S 处的一个或多个 RNA 应该是 18S 和 28S rRNA 的前体。1970 年，Robert Weinberg 和 Sheldon Penman 用 [^{3}H] 甲硫氨酸和 [^{32}P] 标记的脊髓灰质炎病毒侵染 HeLa 细胞，凝胶电泳分离标记的 RNA，发现了关键的中间产物。由于加工过程中的中间产物寿命太短得不到积累，所以检测不到其浓度，但是脊髓灰质炎病毒侵染可减慢加工过程，使中间产物得以检测出来。观察到的主要片段是 45S、41S、32S、28S、29S 和 18S（图 16.3）。前体 rRNA 在真核生物中被甲基化，所以双标记是可行的。

1973 年，Peter Wellauer 和 Igor Dawid 运用电子显微镜观察到人类 rRNA 的前体、中间体和加工产物。发现每种 RNA 都具有自己的分子内碱基配对能力，都有自己的二级结构。Dawid 和 Wellauer 鉴定了所有 RNA 种类（RNA species）的这一“特

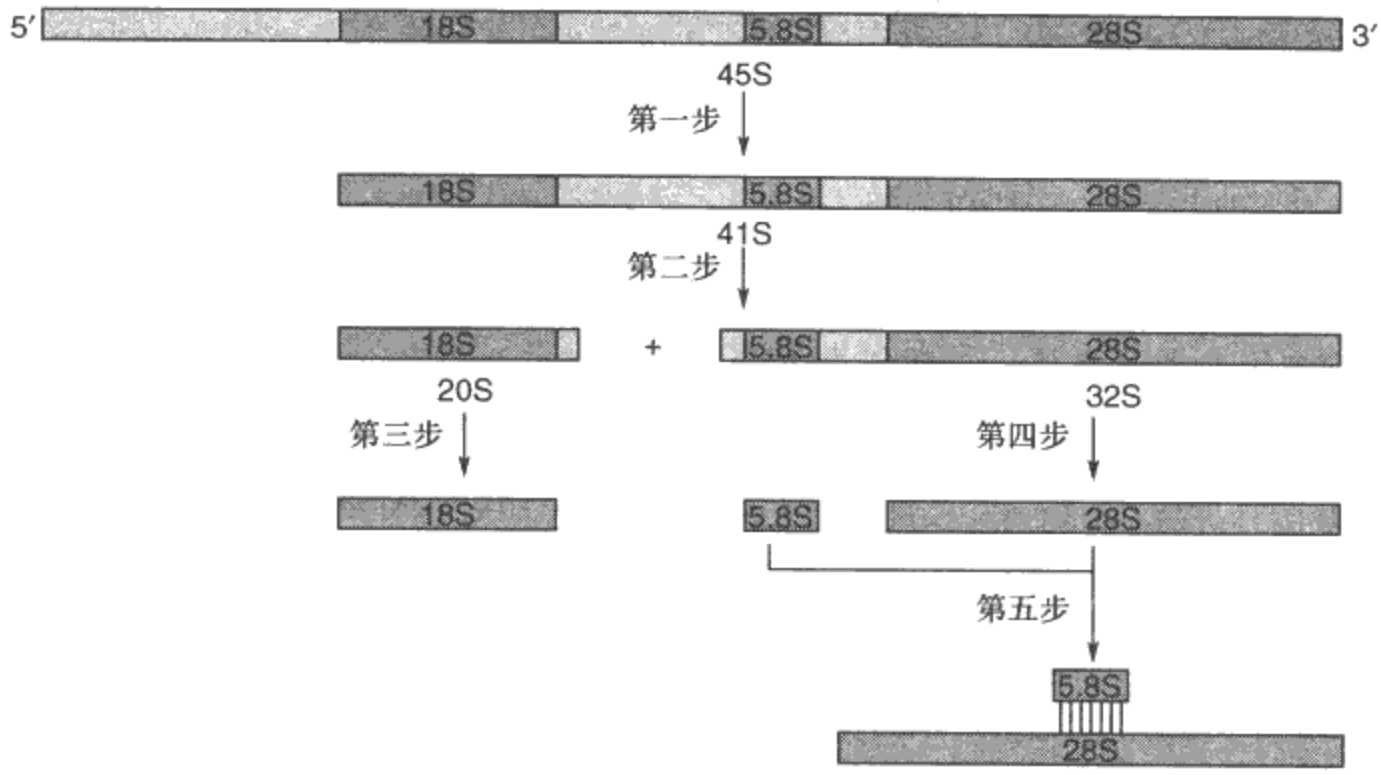


图 16.2　人（HeLa）rRNA 前体 45S 的加工。第一步：45S 前体 RNA 的 5′端被切除，产生 41S 前体；第二步：41S 前体被剪切成两部分，18S rRNA 的前体 20S 和 5.8S、28S rRNA 的前体 rRNA 32S；第三步：20S 前体的 3′端被切除，产生成熟的 18S rRNA；第四步：32S 前体被剪切释放出 5.8S 和 28S rRNA；第五步：5.8S 和 28S rRNA 通过碱基配对连接起来。

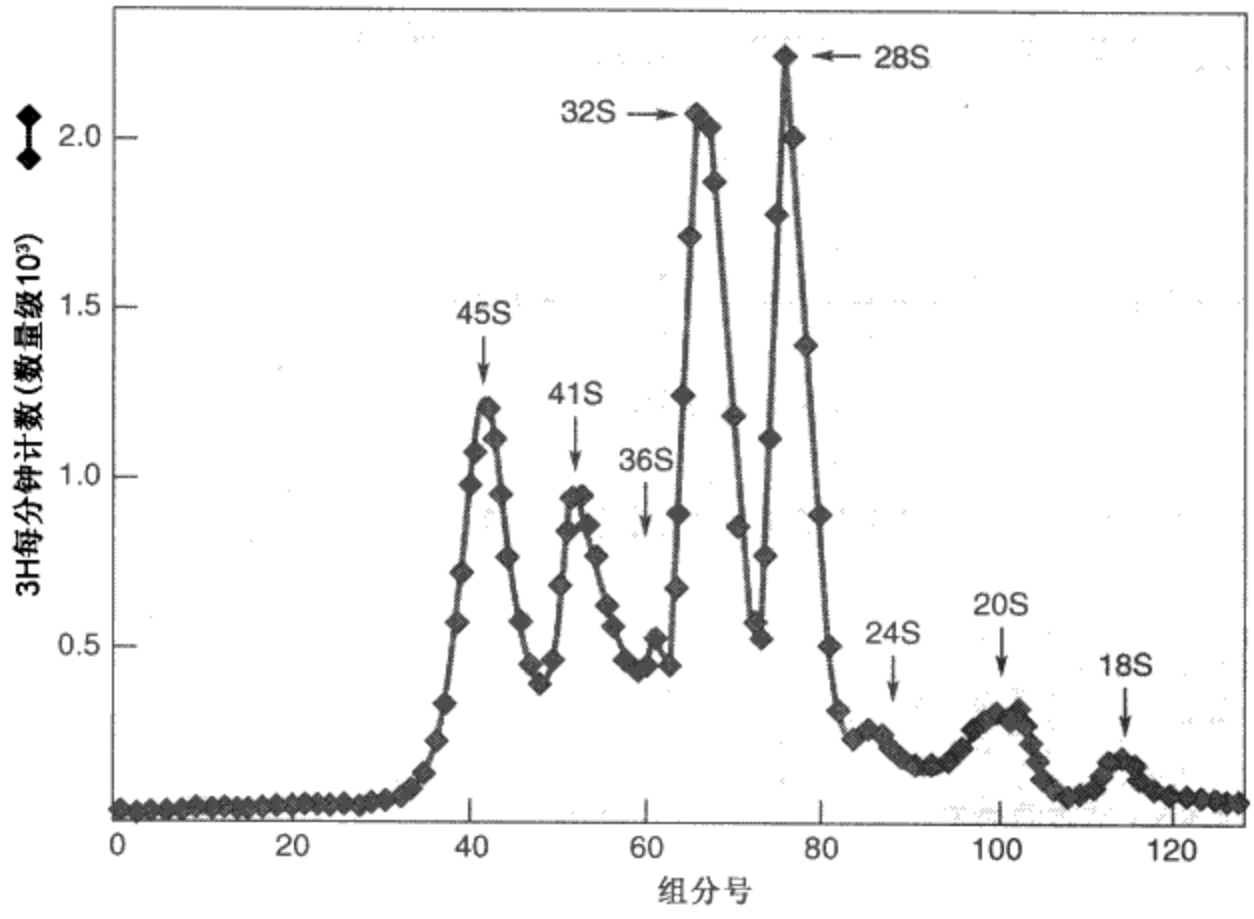


图 16.3　从侵染脊髓灰质炎病毒的 HeLa 细胞中分离 45S rRNA 加工的中间产物。Penman 及同事用［^{32}P］和［^{3}H］甲硫氨酸标记侵染病毒的 HeLa 细胞中的 RNA，使 rRNA 及其前体的许多甲基被标记。从这些细胞中分离出核 RNA（大部分是 rRNA），通过凝胶电泳、切胶，测定每个胶块的^{32}P 和^{3}H 的放射性，然后以放射性强度（cpm 为单位）与切胶组分作图。RNA 种类的迁移率同已知沉淀系数的标准 RNA 相比较。（*Source*：Adapted from Weinberg，R. A. and S. Penman，Processing of 45S nucleolar RNA. *Journal of Molecular Biology* 47：169（1970）.）

征”，从而能在 45S 前体中识别它们，并将 28S 和 18S RNA 定位在 45S 前体中。虽然开始得到的 28S 和 18S 在前体内的顺序是相反的，但我们现在知道这些 RNA 的排列是 5′-18S-5.8S-28S-3′。rRNA 加工过程的细节并不通用，即使小鼠中的 RNA 加工方式也与此有些不同，青蛙的前体只有 40S，比 45S 小得多。但整个真核生物界，rRNA 加工的基本机制，包括成熟序列在前体中的顺序基本上是保守的。

rRNA 的加工步骤是通过一类**核仁小 RNA**（small nucleolar RNA，**snoRNA**）与很多**核仁小核蛋白**（small nucleolar ribonucleo protein，**snoRNP**）的相互作用而被编排好的。细胞中存在成千上万的 snoRNA，但只有少数参加 rRNA 加工过程。已经证明其中 3 个 snoRNP（**E1**、**E2** 和 **E3**）可以与前体 rRNA 的不同区域相互作用。将 E1、E2 和 E3 的反义 RNA 注入非洲爪蟾的卵母细胞，可在不同步骤阻断 rRNA 的加工。这一结果说明 E1 和 E2 在 18S rRNA 的成熟过程中发挥作用，而 E3 则在较早的加工步骤发挥作用，可能参与切割 5.8S rRNA 的上游序列。Joan Steitz 提出 snoRNP 起“分子伴侣”作用，使前体 rRNA 折叠成合适的形状便于加工。

小结 rRNA 在真核生物的细胞核合成，其前体需要加工才能释放成熟的 rRNA。尽管成熟 rRNA 的大小在真核生物之间有所不同，但是所有真核生物的成熟 rRNA 在前体中的排列顺序都是 18S、5.8S、28S。在人类细胞中，rRNA 前体为 45S，经过加工依次产生 41S、32S 和 20S 的中间产物。snoRNA在这些加工过程中起关键作用。

细菌 rRNA 的加工

大肠杆菌有 7 个包含 rRNA 基因的 *rrn* 操纵子。图 16.4a 是其中一个 *rrnD* 操纵子，它除了含有 3 个 rRNA 基因外，还含有 3 个 tRNA 基因。该操纵子可以转录产生一个 30S 前体，经过剪切可释放其中的 3 个 rRNA 和 3 个 tRNA。

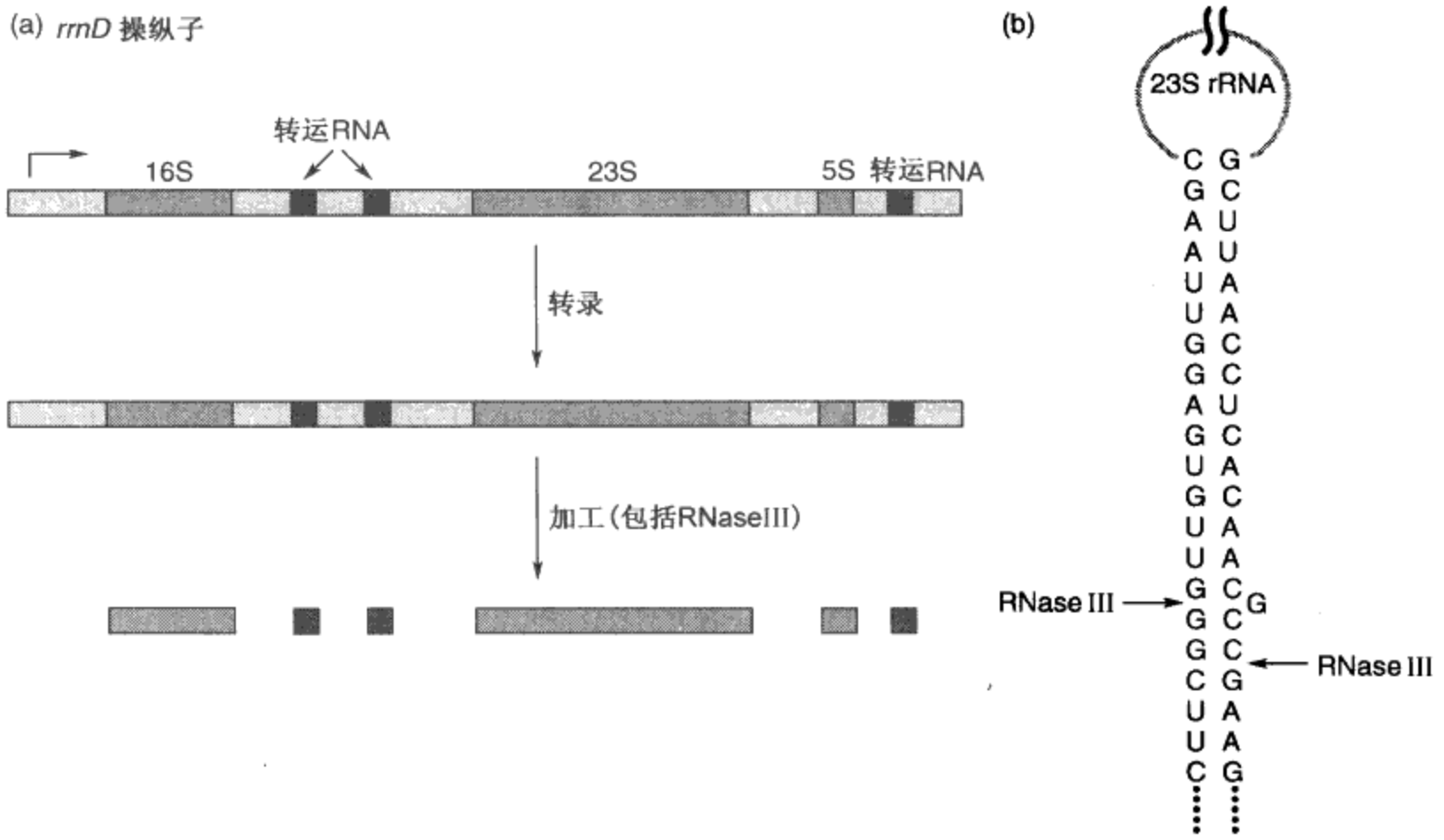


图 16.4 细菌 rRNA 前体的加工。（a）大肠杆菌 *rrnD* 操纵子结构。*rrnD* 操纵子是编码 rRNA 的典型大肠杆菌操纵子，包含 tRNA 的编码区（红色）和 rRNA 的编码区（橙色）。这些编码区都位于转录间隔区（黄色）之间。与通常的细菌操纵子一样，*rrnD* 操纵子转录产生一条长的复合式 RNA，然后由包括 RNase III 在内的 RNA 酶加工产生成熟产物。(b) 序列分析表明，23S rRNA 基因两侧的间隔区互补，可形成延伸的发夹结构，以 23S rRNA 区域为顶部。已观察到的 RNase III 剪切位点位于茎环结构的茎部，偏移 2 个碱基。16S rRNA 基因两侧的区域也能够形成一个更加复杂的茎环结构。

RNase III 至少在起始剪切分离单个大 rRNA 中间体时发挥作用。得出这一结论的遗传学证据是，RNase III 基因缺陷型突变体积累 30S rRNA 前体。1980 年，Joan Steitz 及同事在比较两个不同前体（分别来自 *rrnX* 和 *rrnD* 操纵子）中 rRNA 的间隔区序列时发现它们十分相似。在前体的 16S 和 23S

rRNA 区域两侧都存在互补序列。根据这种互补性，预测有两个伸展的发夹结构，发夹的茎由两个间隔区互补配对产生，两个间隔区之间的 rRNA 成环突出（图 16.4b）。在该模型中，RNase III 的剪切位点在发夹的茎上，另一个核糖核酸酶 **RNase E** 则负责从前体中切除 5S rRNA。

小结 细菌前体 rRNA 含有 tRNA 以及所有 3 个 rRNA。rRNA 是通过 RNase III 和 RNase E 的作用从前体中释放出来的。

16.2 转运 RNA（tRNA）的加工

所有细胞转录出来的 tRNA 都以较长的前体形式存在，必须切除两端多余的 RNA 才能加工为成熟的 tRNA。在真核生物细胞核中，这些前体含有单个 tRNA，而在细菌中，一个前体可能含有一个或者多个 tRNA，有时则同时含有 rRNA 和 tRNA，如图 16.4 所示。由于真核生物和细菌的 tRNA 加工过程非常相似，这里一并讨论。

切开分离多顺反子前体

在含有多个 tRNA 的细菌前体加工中，首先是将前体切成仅含一个 tRNA 的片段。这意味着切割是在前体 tRNA 所包含的各 tRNA 之间，或者在 tRNA 和 rRNA 之间进行的，图 16.4 所示的是既有 tRNA 也有 rRNA 的情况，完成两种 RNA 剪切的酶似乎都是 RNase III。

成熟 5′端的形成

在 RNase III 将前体 tRNA 切割成片段之后，tRNA 仍然含有多余的 5′-和 3′-核苷酸。此时它同真核生物 tRNA 基因的原始转录物很相似，都是含有多余 5′-和 3′-核苷酸的单顺反子前体（单基因）。细菌或真核生物 tRNA 5′端的成熟需要经过一次剪切，剪切位点刚好是成熟 tRNA 的 5′端，如图 16.5 所示。催化此次切割反应的酶是 RNase P。

来自细菌和真核生物细胞核的 RNase P 非常特别。它包含两个亚基，但是不同于我们所研究的其他二聚体酶，它的一个亚基由 RNA 组成，而不是蛋白质。实际上，RNase P 的大部分是由 RNA 组成，因为 RNA（**M1 RNA**）的分子质量约为 125kDa，而蛋白质的分子质量仅为约 14kDa。当 Sidbey Altman 及同事首次分离出 RNase P 并且发现它是一个核糖核蛋白质时，他们面临的关键问题是：哪一部分具有催化活性？RNA 还是蛋白质？当时大

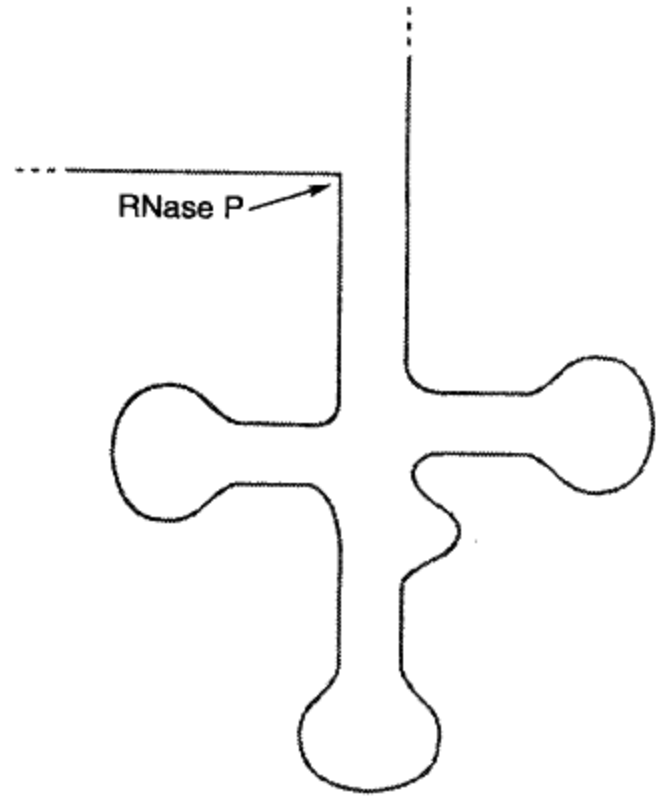


图 16.5 RNase P 的作用。RNase P 的切割位点就是成熟 tRNA 的 5′端。因此，形成成熟 5′端只需要该酶。

多数观点都认为是蛋白质，因为已研究过的酶都是由蛋白质组成的，而不是 RNA。实际上，RNase P 的早期研究表明，当 RNA 和蛋白质分开时，酶便丧失了所有活性。

后来，Thomas Cech 及同事在 1982 年发现了自我剪接内含子存在自主催化活性（第 14 章）。不久，Altman 和 Norman Pace 及同事在 1983 也证实了 RNase P M1 RNA 的催化活性。如图 16.6 所示，调控该酶活性的关键是镁离子浓度。早期的实验使用了 5～10mmol/L 的 Mg^{2+}，在这种条件下，蛋白质和 RNA 两部分都存在时 RNase P 才具有酶活性。图 16.6 显示了 Mg^{2+} 为 5～50mmol/L 浓度时只用 M1 RNA 的效果。他们使用了大肠杆菌 $tRNA^{Tyr}$ 前体和 4.5S RNA 前体的两个不同底物。泳道 1～3 显示在 5mmol/L、10mmol/L、20mmol/L Mg^{2+} 条件下存在差异。在 5mmol/L Mg^{2+} 条件下，两个底物都没有被剪切加工。在 Mg^{2+} 浓度为 10mmol/L 时，仍无法检测到前体 tRNA 的剪切。但加入 20mmol/L Mg^{2+} 时，大约一半前体 tRNA 被剪切为成熟形式，并以单片段形式（图中标记为 5′-Tyr）释放多余的核苷酸。增加 Mg^{2+} 浓度至 30mmol/L、40mmol/L、50mmol/L（分别为泳道 5、7、9）进一步加强了 M1 RNA 对 tRNA 前体 5′端的加工，但是没有引起 4.5S RNA 前体的任何加工。泳道 12 显示原始 RNase P（二聚体形式，含有 RNA 和蛋白质两个亚基）在 10mmol/L Mg^{2+} 条件下对两种底物都可以剪切。

真核生物的核 RNase P 与细菌的 RNase P 非常

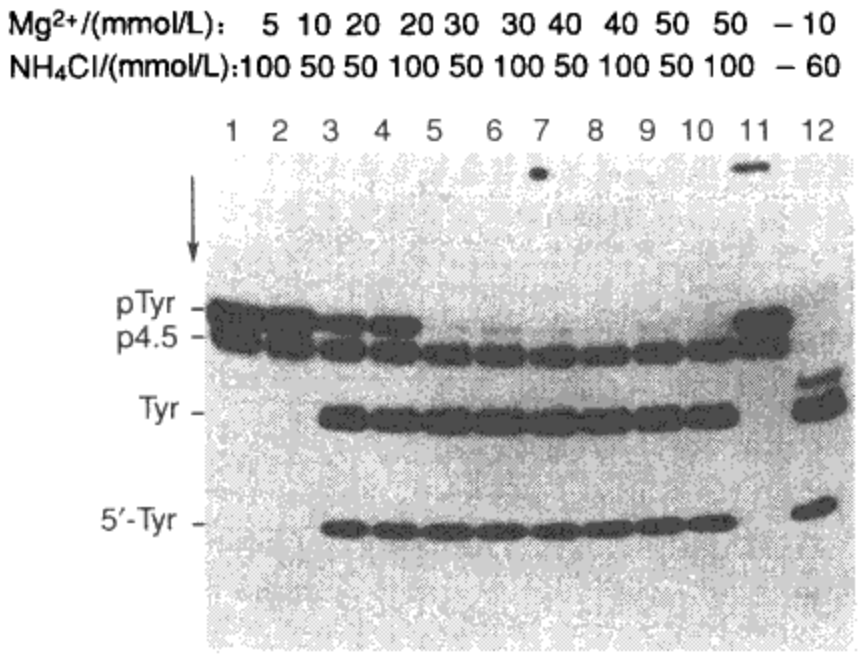


图 16.6　大肠杆菌 RNase P 的 M1 RNA 具有酶活性。Altman 和 Pace 及同事从 RNase P 中纯化出 M1 RNA，在如图所示的 Mg^{2+} 和 NH_4Cl 的浓度下，分别与 ^{32}P 标记的大肠杆菌前体 $tRNA^{tyr}$（pTyr）和 p4.5s RNA 孵育 15min，然后电泳放射自显影观察。泳道 11，无任何添加物。泳道 12，RNase P 的粗提物。在较高 Mg^{2+} 浓度下，M1 RNA 单独就可剪切 pTyr 为成熟的 5′端，但在所有这些条件下，对 p4.5 底物不起作用。（*Source*：Guerrier-Takada，C.，K. Gardiner，T. Marsh，N. Pace，and S. Altman，The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* 35（Dec 1983）p. 851，f. 4A. Reprinted by permission of Elsevier Science.）

相似。例如，酵母细胞核 RNase P 包含蛋白质和 RNA 成分，并且 RNA 具有催化活性。但是 Peter Gegenheimer 及同事却发现菠菜叶绿体 RNase P 好像完全没有 RNA，该酶不受微球菌核酸酶的抑制，因为如果含有催化活性的 RNA，它就受到微球菌核酸酶的抑制。并且，它具有纯蛋白质的密度，而不是主要为 RNA 的核糖核蛋白的密度。人们一直觉得菠菜叶绿体 RNase P 应该含有一个非常小的 RNA，但所有试图检测这种小 RNA 存在的实验都失败了。

小结　前体 tRNA 5′端的多余核苷酸通过 RNase P 的内切核酸酶活性（endonucleolytic）剪切而一步去除。细菌和真核生物细胞核 RNase P 都含有一个具有催化活性的 RNA 亚基，称作 M1 RNA。菠菜叶绿体 RNase P 好像缺少这种 RNA 亚基。

成熟 3′端的形成

tRNA 的 3′端成熟比其 5′端成熟要复杂得多，因为不只是 1 个而是 6 个 RNase 参与其中。Murray Deutscher 和其他研究者已经证明在体外实验中，下列 RNase 参与剪除 tRNA 3′端的核苷酸：RNase D、RNase BN、RNase T、RNase PH、RNase II 以及多聚核苷酸磷酸化酶（polynucleotide phosphorylase，PNPase）。Deutscher 及同事的遗传学实验证明，每个酶对最有效的 3′端加工都是必需的。如果任何一个编码这些酶的基因失活，tRNA 加工的效率都会受到影响，一次性失活所有这些基因对细菌是致死性的。尽管加工效率取决于 RNase 的整体活性，但只要这些酶中存在任何一个都足以保证细胞存活和 tRNA 的成熟。

遗传学和生物化学综合实验显示，RNase II 和 PNPase 共同作用除去 pre-tRNA 的 3′-非转录尾区（3′-trailer）的多余部分。这为 RNase PH 和 RNase T 切除最后两个核苷酸而完成加工过程扫清了道路。RNase T 在切除最后一个核苷酸的反应中最为活跃（图 16.7）。

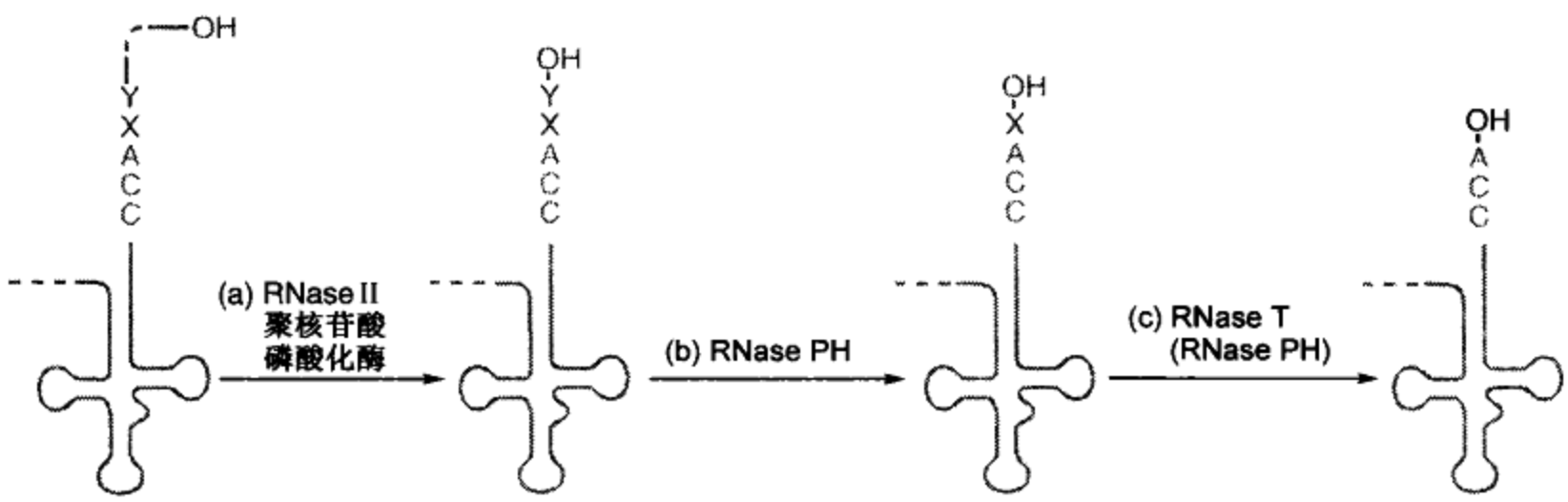


图 16.7　大肠杆菌 tRNA 3′端的加工模式。（a）RNase II 和 PNPase 共同作用，除去 tRNA 前体 3′端的多余碱基，但留下最后 2 个多余碱基（红色）；（b）RNase PH 除去紧接着最后一个碱基的多余碱基（Y，红色）；（c）RNase T 可能在 RNase PH 的帮助下除去最后一个多余碱基（X，红色），留下成熟的 tRNA。

小结 RNase II 和多聚核苷酸磷酸化酶共同作用切除 tRNA 前体中 3′端大部分多余的核苷酸，但停止在+2 阶段，仍留下 2 个多余核苷酸。RNase PH 和 RNase T 在切除最后 2 个核苷酸时最为活跃，并且 RNase T 主要参与移除最后一个多余的核苷酸。

16.3 反式剪接

在第 14 章我们讨论了在几乎所有真核生物种类中都发生的剪接方式，这种剪接方式可以称作**顺式剪接**（*cis*-splicing），因为所涉及的是存在于同一基因中的两个或多个外显子。另一种称为**反式剪接**（*trans*-splicing）的方式中，外显子根本不在同一基因中，甚至不在同一条染色体上，这看起来似乎不可能，但却在一些生物中存在着。

反式剪接机制

反式剪接发生在几种生物中（图 16.8），包括营寄生和自主生活的蠕虫［如秀丽隐杆线虫（*Caenorhabditis elegans*）］，这一现象最早在**锥虫**（trypanosome）中发现。锥虫是一种营寄生带鞭毛的原生动物，能够导致非洲瞌睡病。若基于在本书前面所学的知识，很难对锥虫基因的表达方式做出预测。Piet Borst 及同事在 1982 对编码锥虫表面包被蛋白的 mRNA 的 5′端和编码同一个蛋白质的基因 5′端进行测序时，发现它们并不匹配，这为后来的惊人发现奠定了基础。这个 mRNA 含有 35 个额外核苷酸，是基因中不存在的。随着分子生物学家测序了越来越多的锥虫 mRNA，发现它们都含有相同的 35nt 前导区，称为**剪接前导区**（spliced leader，SL），但是没有任何一个基因编码这种 SL 序列。实际上，SL 由一个单一基因所编码，这个基因在锥虫基因组中重复了 200 多次，只编码 SL，外加一个 100nt 的序列。这个 100nt 的序列通过一个保守的 5′端剪接序列同 SL 相连。所以，这个小基因是由一个短的 SL 外显子组成，并且看起来像个内含子的 5′端。

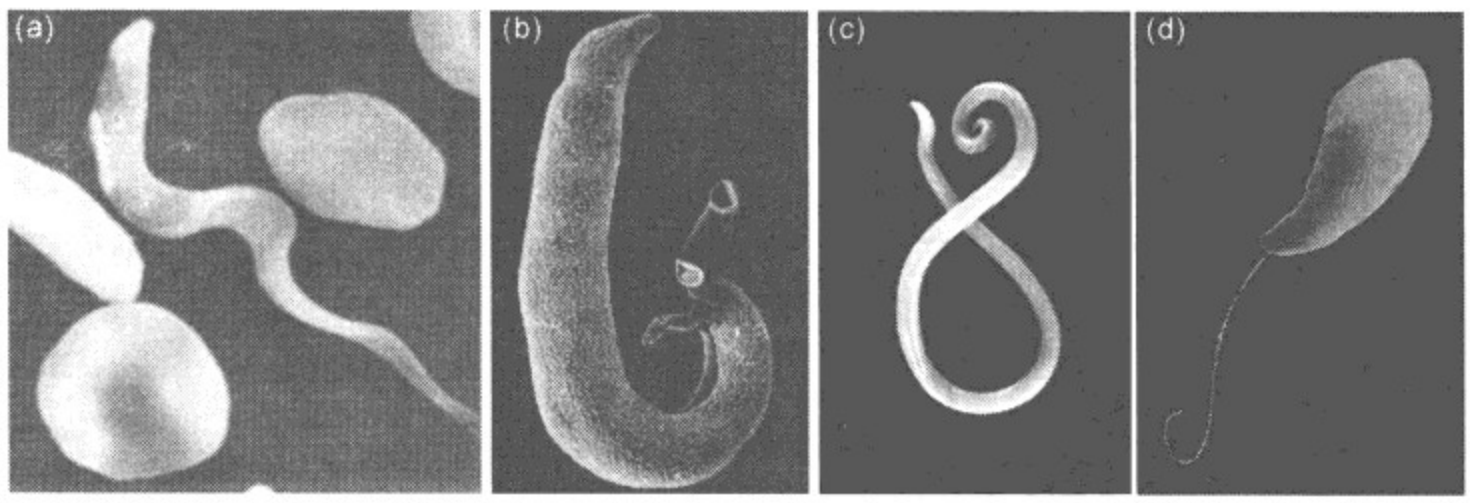
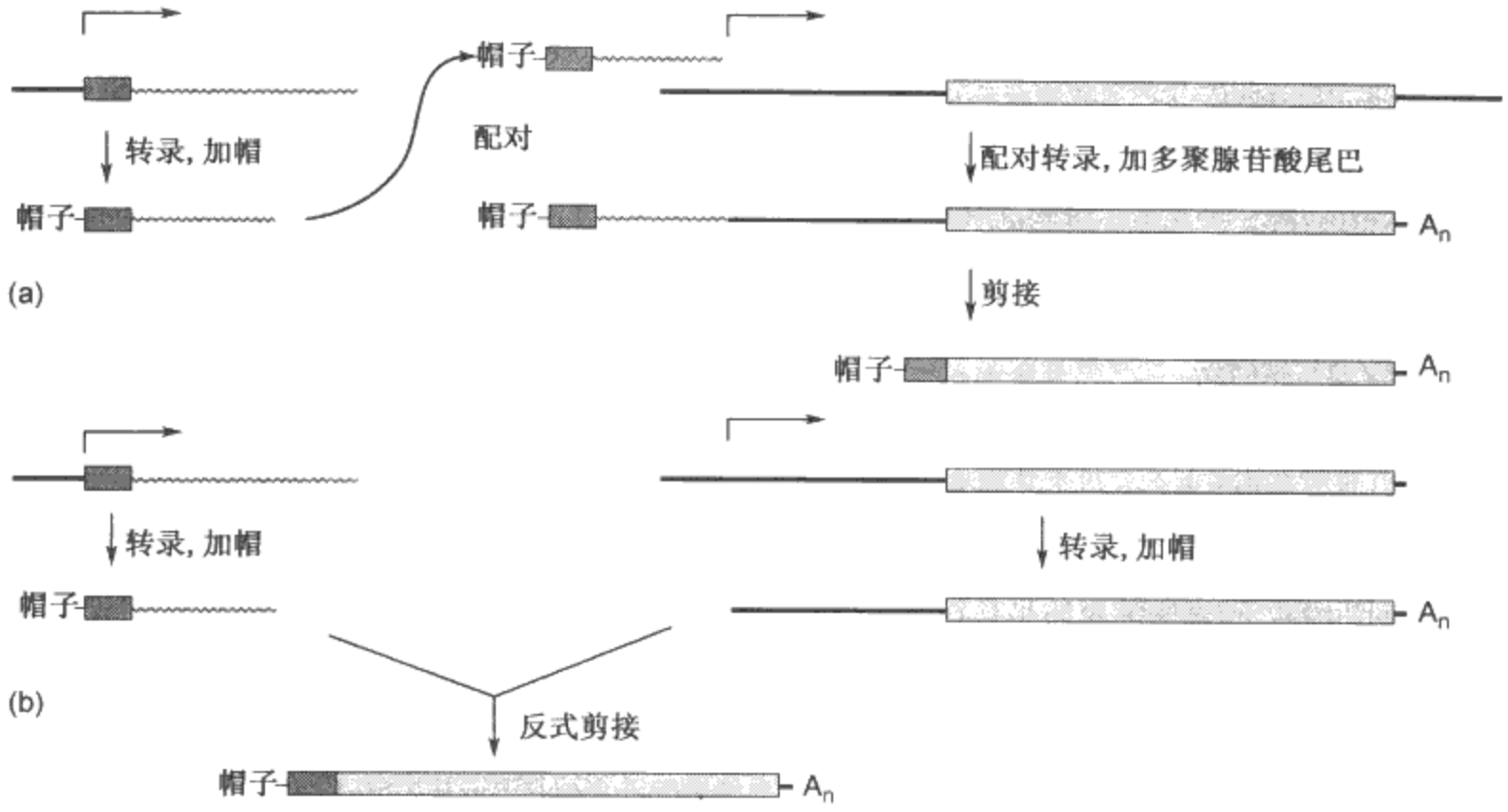


图 16.8 反式剪接的一些生物体。（a）红细胞中的锥虫，由锥虫单个鞭毛所引起的波动，向下延展到细胞的一边；（b）曼森血吸虫（*Schistosoma mansoni*），寄生蠕虫；（c）似蚓蛔线虫（*Ascaris lumbricoides*），另一种寄生蠕虫。(d）眼虫（*Euglena*），一种自由生活的原生生物。（*Sources*：（a）Donelson J. E. and M. J. Turner，How the trypanosome changes its coat. *Scientific American* 45（Feb 1985）.（b）© G. Shih-R. Kessel/Visuals Unlimited.（c）© Arthur M. Siegelman/Visuals Unlimited.（d）© David M. Phillips/Visuals Unlimited.）

我们如何解释这样一个来源于两个相距较远的，甚至不同染色体上独立 DNA 区域所编码的 mRNA 是怎样产生的呢？有两种较合理的解释：一种解释是 SL（有或者无内含子）能够被转录，并且该转录物可作为基因组中其他任何一个编码区的转录引物（图 16.9a）；另外一种解释是 RNA 聚合酶能够分别转录 SL 和编码区域，然后这两个独立的转录物再被剪接在一起（图 16.9b）。

如果这种反式剪接真的发生，那么不应该检测到套索状中间体，而是检测到 Y 形中间体，如图 16.10 所示。Y 形中间体是在内含子的分支点攻击与 SL 外显子毗连的内含子 5′端时形成的。找到 Y 形中间体就使证明反式剪接的真实性前进了一大步。Nina Agabian 及同事在 1986 年报道了 Y 形中间体存在的证据。



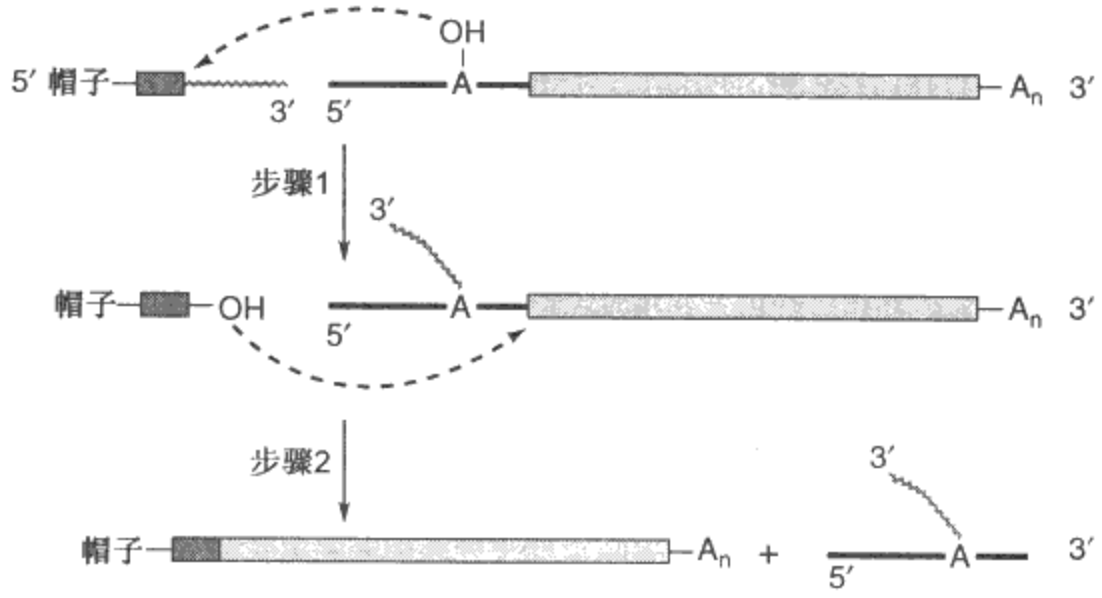


图 16.9　将剪接前导区（SL）连接到 mRNA 编码区的两种假说。（a）由 SL 内含子引发。带有半个内含子（红色）的 SL（蓝色）转录产生 135nt RNA，该 RNA 作为含有半个内含子（黑色）的编码区（黄色）转录的引物，产生一个包括 SL 和编码区以及它们之间整个内含子在内的转录物。这个内含子被剪切产生成熟的 mRNA。(b) 反式剪接。带有半个内含子的 SL 和带有半个内含子的编码区分别转录，然后这两个分离的 RNA 经过反式剪接产生成熟的 mRNA。

图 16.10　锥虫 mRNA 反式剪接的详细过程。步骤 1：连接编码外显子（黄色）的半个内含子（黑色）中的分支点腺苷酸攻击前导外显子（蓝色）与其半个内含子（红色）之间的连接点，产生 Y 形内含子-外显子中间体，类似于顺式剪接所产生的套索中间体。步骤 2：前导外显子攻击分支内含子和编码外显子之间的剪接位点，产生成熟的 mRNA 和 Y 形内含子。

Y 形结构区别于一般套索状中间体的特征是，在 Y 形结构中 SL 内含子的 3′端是游离的（图 16.10）。这表明用能断裂 2′-5′-磷酸二酯键的去分支酶（debranching enzyme）处理 Y 形中间体，应该产生一个 100nt 的副产物（图 16.11），这与套索中间体仅被简单地线性化的结果相反。图 16.12 表示去分支酶处理后，以 100nt 特异性核苷酸片段为探针，对总 RNA 和有 poly（A）尾的 RNA 进行 Northern 杂交的结果。在这两个样本中，都出现了 100nt 的片段，从而证实了反式剪接假说。

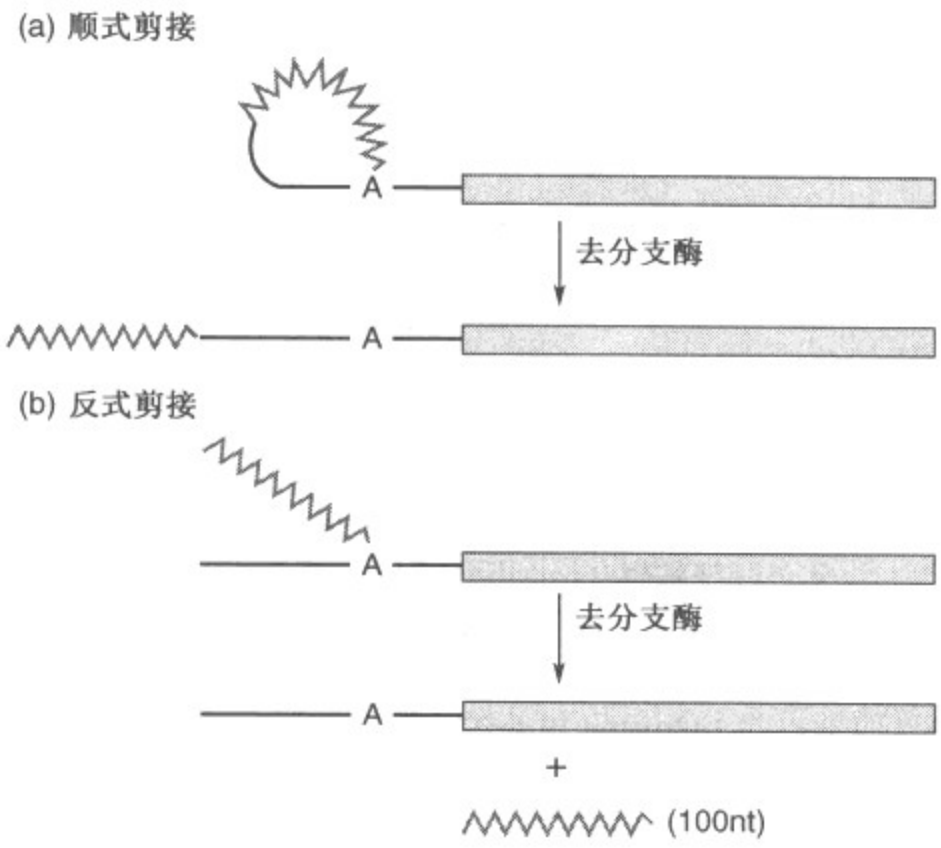


图 16.11　用去分支酶处理假设的剪接中间体。（a）顺式剪接。去分支酶只是打开套索使之呈线形；（b）反式剪接；100nt 的半个内含子（红色）在 3′端是开放的而不是套索结构，去分支酶将它作为独立 RNA 释放。

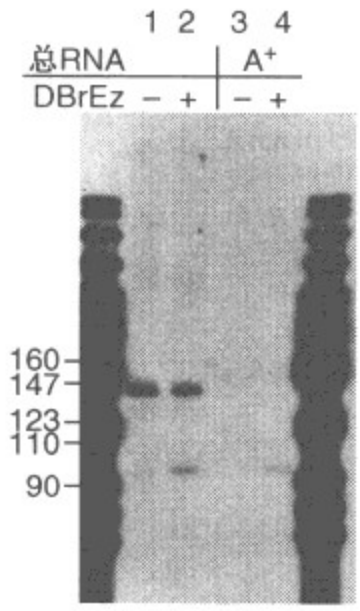


图 16.12　用去分支酶从较大的 RNA 上释放 SL 的半个内含子。如图顶部所示，Agabian 及同事用^{32}P 标记锥虫的 RNA，用去分支酶（DBrEZ）处理总 RNA 或 poly（A）$^{+}$RNA，电泳并印迹产物，以 100nt SL 内含子特异性寡聚核苷酸为探针检测印迹膜，在酶处理的两种 RNA 样品中都可以清楚地检测到探针信号。（*Source*：Murphy W. J.，K. P. Watkins，and N. Agabian，Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing. Evidence of *trans*-splicing. *Cell* 47（21 Nov 1996）p. 521，f. 5. Reprinted by permission of Elsevier Science.）

小结　锥虫 mRNA 是通过反式剪接短引导区外显子和任何一个独立编码外显子而形成的。

16.4　RNA 编辑

反式剪接并不是锥虫唯一的特殊 RNA 加工形式。这种生物含有不寻常的线粒体，称作**动基体**（kinetoplast），其中含有两类环状 DNA，它们连接在一起形成网络状（图 16.13）。动基体有 25～50 个相同的**大环**（maxicircle），大小为 20～40kb，含有线粒体基因。另外还有大约 10 000 个 1～3kb 的**小环**（minicircle），在线粒体基因表达中起调节作用。1986 年，Rob Benne 及同事发现，锥虫的**细胞色素氧化酶**（cytochrome oxidase，COII）mRNA 序列与 COII 基因序列不匹配，mRNA 中含有 4 个在基因中

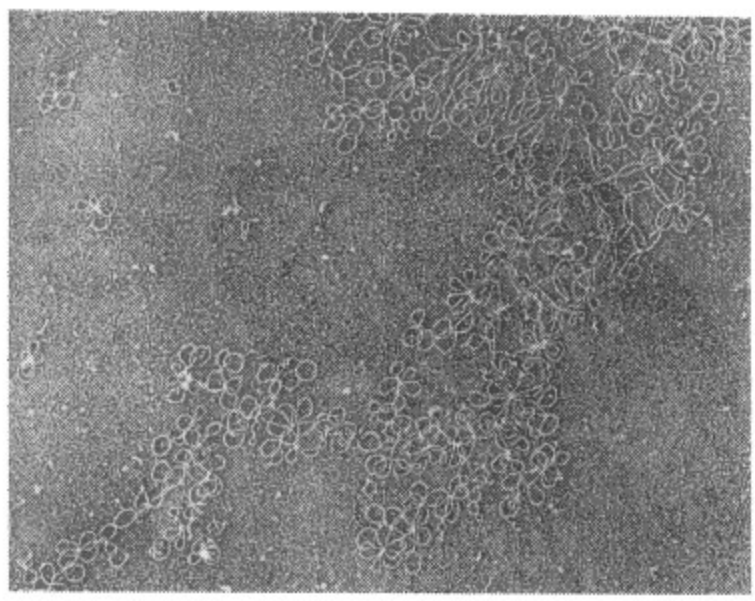


图 16.13　利什曼原虫动基体小环和大环的部分网络图片。（*Source*：*Cell* 61（1 June 1990）cover（acc. Sturm & Simpson，pp. 871-84）. Reprinted by permission of Elsevier Science.）

缺失的核苷酸（图 16.14）。这些缺失的核苷酸可能引起**移码**（frameshift，在核糖体阅读 mRNA 时读码框发生移位，见第 18 章），进而导致基因失活。但锥虫以某种方式给 mRNA 提供了这 4 个核苷酸，避免了移码。

COX II DNA: ···GTATAAAAGTAGA　G A ACCTGG···

COX II RNA: ···GUAUAAAAGUAGAUUGUAUACCUGG···

图 16.14　锥虫的 COII 基因的部分序列与它的 mRNA 产物的比较。mRNA 中的 4 个 U 并不表示基因中的 T，这 4 个 U 大概是通过编辑加入到 mRNA 中的。

当然，不排除 Benne 及同事测序的那个基因是不编码 mRNA 的**假基因**（pseudogene）的可能性。假基因是个发生突变后没有功能、不再使用的基因拷贝。活性基因可能位于其他地方，只是研究人员可能没有找到它们。但这种解释的问题是，Benne 及同事尽可能地寻找，仍没有在动基体或核中找到其他 COII 基因。而且，他们在另外两种锥虫中却发现了 COII 基因存在同样的核苷酸缺失。据此，Benne 及同事推断，锥虫的 mRNA 来源于称为**隐秘性基因**（cryptogene）的不完整基因，然后通过添加缺失的核苷酸（全是 UMP）进行编辑。

到 1988 年，大量锥虫动基体基因和相应的 mRNA被测序。结果显示，在这些生物中 RNA 编辑是一种常见现象。实际上，有些 RNA 进行了**广泛编辑**（panedited）。例如，布氏锥虫（*Trypanosoma brucei*）COIII mRNA 含有一个 731nt 延长序列，其中的 407 个 UMP 是通过编辑加入的，这段延长序列中的 19 个编码的 UMP 在编辑时被删除了。部分序列如图 16.15 所示。

图 16.15　布氏锥虫 COIII mRNA 部分被编辑的序列。通过编辑加入的 U 以灰色的小写字母表示，基因中出现的 T 但在 mRNA 中却不出现（如 U）的以序列上方的蓝色表示。（*Source*：Adapted from *Cell* 53：cover，1988.）

RNA 编辑的机制

我们已经假定编辑是一种转录后加工事件。这看起来像一个好赌注，因为同一 mRNA 的未编辑转录物和编辑后的版本可以同时出现。而且，编辑发生在 mRNA 的 poly（A）上，poly（A）是转录后加入的。

编辑机制的一个重要线索是部分编辑的转录物已经分离得到，它们常常发生在 3′端而不是 5′端。这强烈提示 RNA 编辑是按 3′→5′方向进行的。1988 年，Kenneth Stuart 及同事首次报道了这一现象。他们利用 RT-PCR 技术，先以 RNA 为模板在反转录酶的作用下产生第一条 DNA 链，然后进行标准的 PCR（第 4 章）。

在实验中，他们用了几组不同的 PCR 引物对：两条编辑的引物、两条未编辑的引物、仅一条编辑过的引物。完全编辑的 RNA 只能与编辑的引物杂交，产生 PCR 信号，而不能与非编辑的引物杂交，

所以，只有一条未编辑引物的 PCR 反应均不会给出该 RNA 的 PCR 信号。相反，一个完全未编辑的 RNA 只能与未编辑引物发生反应。他们在实际实验中，利用未编辑的 5′引物和编辑的 3′引物检测 3′端发生编辑的转录物，或利用编辑的 5′引物和未编辑的 3′引物检测 5′端发生编辑的转录物。如果转录物编辑的方向是 3′→5′，那么只能在 3′端编辑的转录物中检测到 PCR 信号，而 5′端编辑的转录物检测不到 PCR 信号。该 PCR 方法的优点是可以扩增量非常小的 RNA，如部分编辑的 RNA，将其变为易于检测的 DNA 条带。

图 16.16 是分析结果。泳道 1～4 是布氏锥虫动基体 RNA 与不同引物组合的 PCR 产物。只有两条引物都编辑时，或 3′引物编辑时，才能看到 PCR 信号。仅有 5′引物编辑时，看不到任何信号。因此，3′编辑发生在 5′编辑缺失时，但是在 3′编辑未发生时，5′编辑就不能发生。这与编辑按 3′→5′方向进行的推论是一致的。泳道 5～6 和泳道 7～10 分别表示阳性对照和阴性对照。

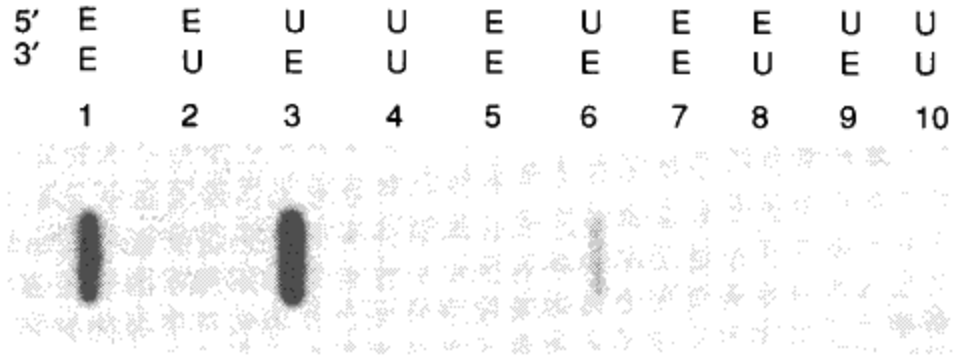


图 16.16　编辑方向的 PCR 分析。如上所示，Stuart 及同事利用动基体 RNA 和编辑的（E）或未编辑的（U）5′引物与 3′引物进行 RT-PCR，检测细胞色素 c 氧化酶 III 的转录情况，如图上方所示。然后印迹转移 PCR 产物并与标记的探针杂交，放射性自显影检测杂交产物。PCR 模板：泳道 1～4，野生型细胞的 RNA；泳道 5 和 6，3′编辑的 cDNA（阳性对照）；泳道 7～10，缺少线粒体 DNA 的突变体 RNA（阴性对照）。(*Source*: Abraham, J. M., J. E. Feagin, and K. Stuart, Characterization of cytochrome *c* oxidase lll transcripts that are edited only in the 3′ region. *Cell* 55 (21 Oct 1988) p. 269, f. 2a. Reprinted by permission of Elsevier Science.)

实验结果很有价值，但有个缺陷，未编辑 3′引物的泳道都没有 PCR 信号。我们期望在泳道 4 中看到信号，它用的是未编辑 5′引物和 3′引物，但是什么也没有看到。这可能（也许）是总的未编辑 RNA 的浓度太低，用这个方法无法检测到。但也可能是 3′未编辑引物存在问题。所以，该实验可以增加一个 3′未编辑引物的阳性对照，以便消除对 3′未编辑引物质量的怀疑，可以用该基因的体外转录物做模板，它是完全没有编辑的，如果检测到 PCR 信号，表明 3′未编辑引物质量没问题。对照在 PCR 实验中尤为重要，因为 PCR 对少量核苷酸，包括污染的核苷酸也有很强的扩增能力。

什么决定了编辑系统应该在哪里加入或删除 UMP 呢？Larry Simpson 及同事在 1990 年发现利什曼原虫（*Leishmania*）大环中编码的**引导 RNA**（guide RNA，**gRNA**），由此获得了答案。他们从计算机检索当时已知的大环 DNA 序列的 21kb 部分开始。检索结果揭示了 7 个短序列可以产生短 RNA（gRNA）与 5 个不同编辑的线粒体 mRNA 部分互补。这种 gRNA 能够在 mRNA 的几十个核苷酸范围内指导 UMP 的插入或删除，如图 16.17a 和图 16.17b 所示。一旦这一编辑完成，另一个 gRNA 就与刚编辑区的 5′端杂交，指导新一段的编辑，如图 16.17c 和图 16.17d 所示。以这种方式从 mRNA 的 3′端向 5′端方向，后续 gRNA 结合到前任 gRNA 编辑过的区域，指导进一步编辑直到完成整个编辑过程。gRNA 的序列再次证实了编辑方向是 3′→5′的结论：只有在编辑的 3′边界的 gRNA 能与未编辑序列杂交。所有其余的 gRNA 与编辑过的序列杂交，这只有在编辑从 3′→5′方向进行时才合理。

gRNA 和 mRNA 碱基配对的一个显著特点是存在 G-U 碱基配对和标准的 Watson-Crick 碱基配对。在第 18 章，我们将学习 G-U 碱基配对在翻译的密码子-反密码子配对，其中一个碱基通过轻微摆动在 Watson-Crick 碱基对中的位置，可适应这种非标准碱基配对。G-U 碱基配对在编辑中的重要性可能是因为它们较 Watson-Crick 碱基对弱。这表明新 gRNA 的 5′端可以取代前一个 gRNA 的 3′端，因为它可与 mRNA 新编辑区形成 Watson-Crick 碱基配对，而前一个 gRNA 的 3′端与 mRNA 的配对中有弱的 G-U 配对（图 16.18）。

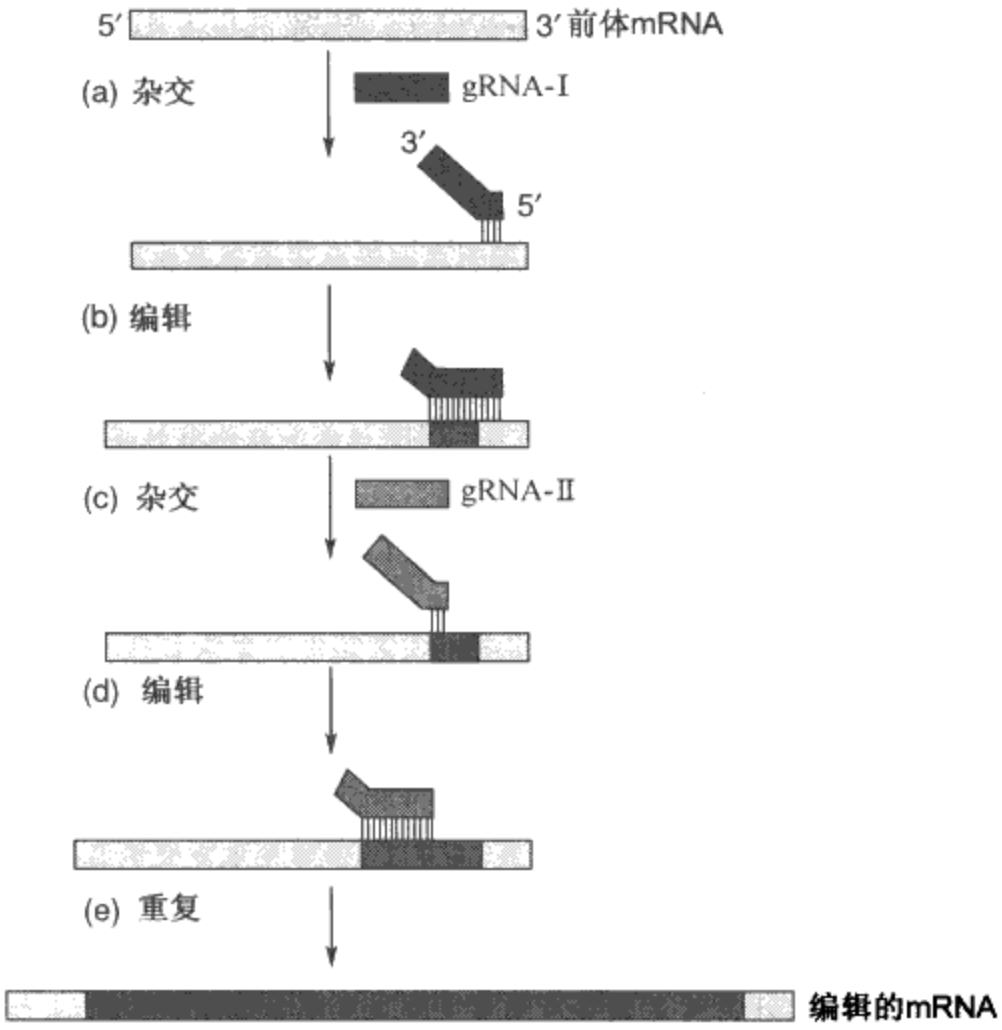


图 16.17　gRNA 在编辑中的作用模式。（a）第一步，gRNA-I（蓝黑色）的 5′端与前体 mRNA 的无须编辑部分杂交，其 3′端通过 oligo（U）区域与前体 RNA 杂交，但图中未显示；（b）其余的大部分 gRNA-I 指导部分前体 mRNA 的编辑，编辑过的部分用红色表示，由于插入了 UMP 前体 mRNA 长度增加；（c）一个新的 gRNA，即 gRNA-II（浅蓝色）通过与前体 mRNA 刚编辑区的 5′端杂交而取代了 gRNA-I；（d）gRNA-II 指导新一部分前体 mRNA 的编辑；（e）以下一个 gRNA 重复前面的步骤，直到 RNA 完全编辑。

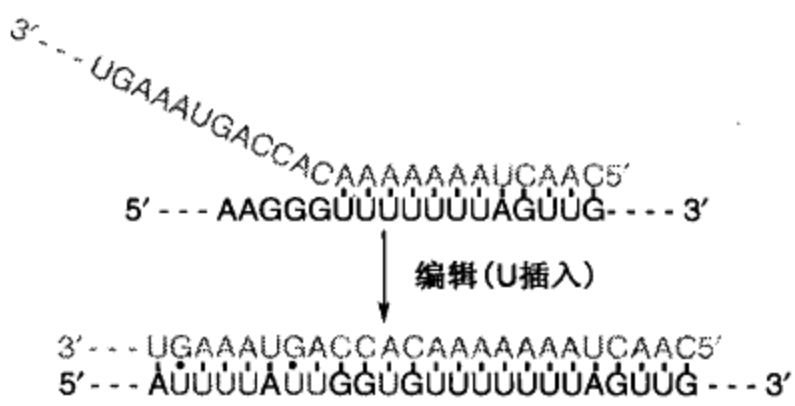


图 16.18　假设 RNA 的一部分的编辑。gRNA（蓝色）通过 Watson-Crick 碱基配对与前体 mRNA 的已编辑部分结合，然后 gRNA 的 3′端作为 U（红色）插入的模板。新插入的 U 和 gRNA 之间的碱基配对大部分是 Watson-Crick 的 A-U 配对，但有 2 个是摇摆的 G-U 配对，用圆点表示。

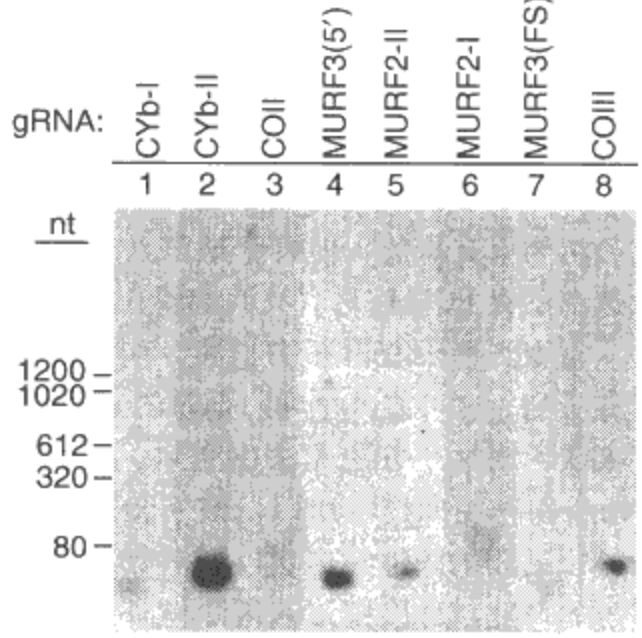


图 16.19　gRNA 存在的证据。Simpson 及同事对利氏曼原虫线粒体 RNA 进行 Northern 杂交，用与 gRNA 杂交的标记寡核苷酸探针检测印迹。检测的 gRNA 在图的上方标出。（*Source*：Blum，B.，N. Bakalara，and L. Simpson，A model for RNA editing in kinetoplastid mitochondria：“Guide” RNA molecules transcribed from maxicircle DNA provide the edited information. *Cell* 60（26 Jan 1990）p. 191，f. 3a. Reprinted by permission of Elsevier Science.）

1990 年末，Nancy Sturm 和 Larry Simpson 发现小环也编码 gRNA。他们除了证实编码 gRNA 的可能性外，还找到了 gRNA 存在的直接证据。将动基体 RNA 进行电泳，Northern 印迹，根据推测的 gRNA 基因在大环中的序列设计寡核苷酸标记探针与其进行杂交，检测 gRNA。图 16.19 显示，这个实验检测到了小 RNA，它们大部分都短于 80nt。

编辑的精确机制及插入和删除 UMP 所需要的切割和连接，一直都不清楚，但在动基体中发现的酶活性为这些问题提供了一些线索。例如，动基体中含有**末端尿苷酰转移酶**（terminal uridylyl transferase，**TUTase**），能够在编辑的过程中将多出的 UMP（尿嘧啶核苷酸）加入到 mRNA 中。因为

mRNA 必须切开才能接受新的 UMP，所以还必须再次连接起来，而动基体也恰好含有 **RNA 连接酶**（RNA ligase）。剩下的主要问题是用于编辑的尿嘧啶的来源，这可由 UTP 提供。此外，在 gRNA 末端的尿嘧啶通过酯交换反应（transesterification）而转移到前体 mRNA 上，即尿嘧啶可从 gRNA 末端脱去，直接转移到前体 mRNA 上。

1994 年，Scott Seiwert 和 Stuart 用线粒体提取物和 gRNA 来编辑一个人工合成的前体 mRNA。他们发现 UMP 的剔除需要 3 个酶的活性（图 16.20a）。①沿着 gRNA 的作用方向在 UMP 需要被移除的地方切开 mRNA 的内切核酸酶；②特异性切割末端尿嘧啶的 3′外切核酸酶；③RNA 连接酶。1996 年，Stuart 及同事用相似的体外系统证明，UMP 的插入遵循以上相似的 3 步途径（图 16.20b）。①gRNA 介导的内切核酸酶在需要插入 UMP 的位点切割 mRNA；②由 gRNA 引导，有一个酶（可能是 TUTase）从 UTP（不是来自于 gRNA 的那个）转移 UMPs；③RNA 连接酶，将两条 RNA 重新连接在一起。有意思的是，gRNA 由线粒体 DNA 编码，而编辑所需要的蛋白质由细胞核编码，而后运输到线粒体。

小结 锥虫线粒体所编码的 mRNA 不完整，翻译之前要进行编辑加工。编辑在一个或者多个连续的 gRNA 的作用下沿着 3′→5′的方向进行。第一个 gRNA 的 5′端杂交到一个前体 mRNA 待编辑位点——3′未编辑区域上，剩下的 gRNA 的 5′端依次杂交到已编辑区域，该区域的 5′区域将要被编辑。所有这些 gRNA 提供了 A 和 G，作为 U 掺入的模板。有时 gRNA 因没有 A 或 G，不能与 mRNA 的 U 配对，这种情况下 mRNA 中的 U 将被移除。移除 U 的机制包括：①在 U 移走之前切割前体 mRNA；②通过外切核酸酶移除 U；③将两条前体 mRNA 连接起来。U 插入机制的第一步和最后一步同上面的机制相同，但中间步骤（第二步）涉及增加一个或者多个 U 而不是移除 U，增加的 U 来源于 UTP。

核苷酸脱氨基化编辑

RNA 编辑不是一个仅仅在奇怪生物中才发生的稀奇事，它在高等生物甚至哺乳动物中也发挥着重要作用。在哺乳动物中虽然没有发现像锥虫那样进行尿嘧啶的插入和删除的编辑，但有大量证据表明存在着另一种方式的编辑，即腺苷酸脱氨基后转化为次黄嘌呤核苷酸（inosine），腺苷酸氨基基团被氧取代。因为次黄嘌呤核苷与鸟嘌呤核苷以相同的方式与胞嘧啶核苷酸碱基互补配对，腺苷酸脱氨基化后改变了一个密码子的编码意义。例如，ACG（苏氨酸）密码形成了 ICG 密码，它将被核糖体作为 GCG（丙氨酸）读码。

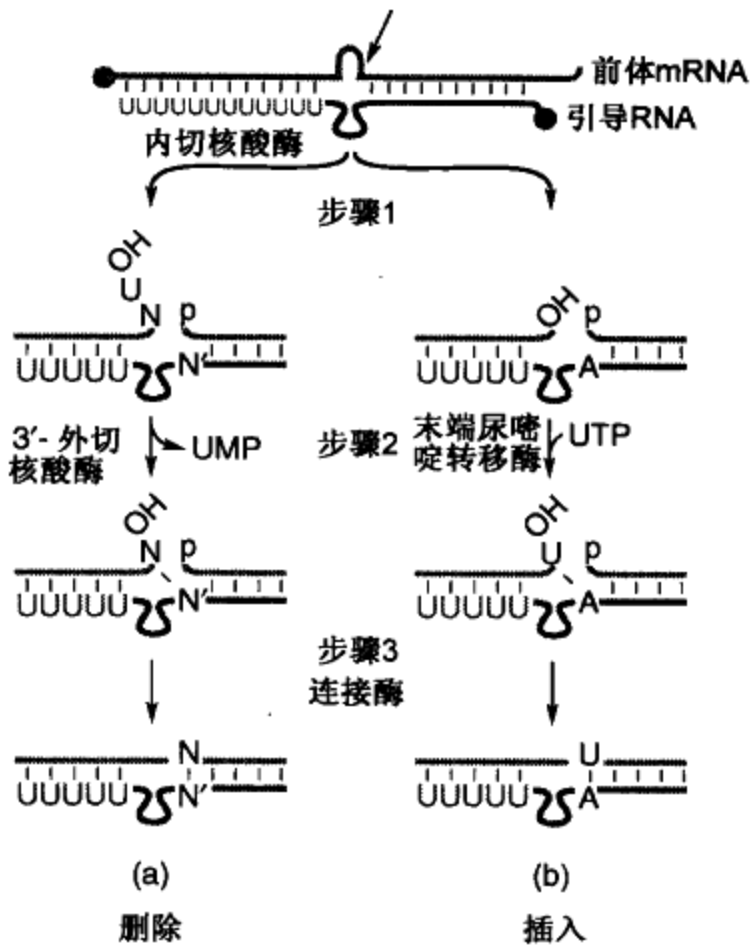


图 16.20 RNA 编辑机制。显示删除 U（a）和插入 U（b）的机制，从 mRNA 前体（紫红色）和 gRNA（蓝黑色）的杂交开始，如图上部所示。gRNA 的突出部分表示与 mRNA 前体不匹配的碱基。箭头表明核酸酶剪切需编辑 mRNA 的位点。(a) 删除 U。步骤 1，内切核酸酶剪切 mRNA 前体中将被删除 U 的 3′端。步骤 2，外切核酸酶切去左手 RNA 片段末端的 UMP，在 mRNA 前体的 N 碱基和 gRNA 的 N′碱基之间进行碱基配对。步骤 3：RNA 连接酶将 mRNA 前体的两部分又连接在一起。(b) 插入 U。第一步，内切核酸酶剪切 mRNA 前体中由 gRNA 指导的 U 将被插入的位置。第二步，TUTase 将 UMP 从 UTP 转移到左手 RNA 片段的 3′端，在 gRNA 中 U 和 A 配对。第三步，RNA 连接酶将 mRNA 前体连接起来。(*Source*：Adapted from Seiwert，S. D.，Pharmacia Biotech in Science Prize. 1996 grand prize winner. RNA editing hints of a remarkable diversity in gene expression pathways. *Science* 274：1637，1996.)

这类 RNA 编辑由 **RNA 依赖性腺嘌呤脱氨酶**（adenosine deaminase acting on RNA，**ADAR**）介导。人和小鼠含有 3 个 ADAR 基因：*ADAR1*、*ADAR2*、*ADAR3*。前两个基因产物在体内普遍存在，但第三个基因的产物只存在于脑中。这些酶非

常特异，如果它们将 mRNA 的每个腺苷都去氨基的话，那将是灾难性的，所以它们只选择特定的mRNA中特定的腺苷发挥作用。例如，ADAR2 将谷氨酸盐敏感型离子通道受体亚基 B（GluR-B）mRNA 的一个腺苷去氨基化，具有大于 99%的效率。mRNA中的这一改变将一个谷氨酸密码换成了精氨酸密码。这一改变重要吗？我们知道含 GluR-B 蛋白的离子通道中的精氨酸被谷氨酸替代后，对钙离子具有较强的通透性。因此可以预测 *ADAR2* 基因缺陷小鼠会有严重问题。的确如此，纯合 *ADAR2* 基因缺陷型小鼠不能进行恰当的 GluR-B mRNA 编辑。它们看似发育正常，但在断奶后很快死亡。

Peter Seeburg 及同事想了解，如果对小鼠 *GluR-B* 基因进行简单的诱变，使之在原编辑位点编码精氨酸，然后无需对该基因的转录物再进行编辑了，那么将会发生什么情况呢？在进行这个实验时，他们发现，即使带有纯和的缺陷型 *ADAR2* 基因的小鼠也是活的。因此，该实验也证明 ADAR2 酶的唯一关键靶位点是 GluR-B 转录物。

果蝇基因组只含有一个 *ADAR* 基因。一旦该基因发生突变，突变果蝇将缺少所有的 ADAR 活性，不能对已知的编辑位点进行任何 mRNA 的编辑。这些突变的果蝇能够存活，但是它们行走困难、不能飞、进行性神经元退化。所以，该突变的表型与哺乳动物中 *ADAR2* 基因突变的表型相似。果蝇实验证明了 ADAR 对 mRNA 的编辑是中枢神经系统正常发育所必需的这一假说。

ADAR1 对于哺乳动物的正常生命活动似乎也是必需的。Kazuko Nishikura 及同事对小鼠干细胞实施诱变，得到杂合的突变体（$ADAR1^{+/-}$），然后将这些细胞注射到正常小鼠的胚泡中，期望产生嵌合鼠（第 5 章）。但他们发现这是不可能的，即使利用大比率的突变细胞也不能产生嵌合鼠。注入多于一定量的突变体细胞的胚胎不会存活到出生。所以，*ADAR1* 突变体是胚致死性的，尽管是杂合的。

为什么具有低 ADAR1 活性的胚会死亡呢？受影响胚胎的大多数组织表现正常，但红细胞却例外。如同来自卵黄囊中的红细胞一样，它们从卵黄囊转移到肝脏后的很长时间内却仍具有细胞核，通常红细胞在肝脏内会失去细胞核。所以，红细胞发生的某些机制依赖于胚胎中 *ADAR1* 的完全表达。

有意思的是，有些肿瘤丧失了 ADAR 活性。尤其是叫做多形性成胶质细胞瘤（gioblastoma multiforme，GBM）的人类恶性脑瘤，仅有非常低的 ADAR2活性和相应的编辑不良的 *GluR-B* mRNA。一些癫痫患者也有编辑不良的 mRNA，而且 GBM 患者常遭受癫疯病的困扰。

另一种编辑方式是通过 **RNA 依赖性胞嘧啶脱氨酶**（cytidine deaminase acting RNA，CDAR）进行的，将胞嘧啶转变成尿嘧啶。这种 C→U 的编辑在多发性神经纤维瘤 I 型患者的良性神经鞘瘤中大约 25%是缺陷型的。C→U 编辑也可能发生在人类细胞的 HIV 转录物中。在被 HIV 感染的人类细胞中仍存在另一种编辑方式，即 G→A。但是这种编辑不能仅用一步去氨基作用来解释，目前仍不清楚它的作用机制。

小结 包括果蝇和哺乳动物在内的高等真核生物，其 RNA 必须经过转录后的编辑加工，腺苷酸脱氨基化形成次黄嘌呤核苷酸才能编码适当的蛋白质。已知 RNA 依赖性腺嘌呤脱氨酶（ADAR）行使这种 RNA 编辑功能。此外，一些腺苷必须脱氨基成为尿苷后 RNA 才能正确编码。

16.5 基因表达的转录后调控

在对真核和原核转录机制的讨论中，我们已经看到很多转录水平调控的例子。通过阻断基因表达的第一步转录来控制基因表达是合理的，这也是最经济的方法，因为细胞不需要消耗能量去合成一个无用蛋白质的 mRNA。

虽然转录水平的调控是控制基因表达的最普遍形式，但绝不是唯一方式。我们在第 15 章知道 poly(A)可以稳定 mRNA 并确保 mRNA 正常翻译，mRNA 3′-UTR 具有一个特殊序列，称为胞质多聚腺苷酸化元件（cytoplasmic polyadenylation element，CPE），它在卵母细胞成熟过程中控制着母性信息的多聚腺苷酸化效率。可见，这些 CPE 行使基因表达控制者的功能。

基因表达的更重要转录后调控是 mRNA 稳定性的调控。Joe Harford 指出“细胞内 mRNA 的水平与其稳定性的相关性较与其转录效率更为密切”。

酪蛋白 mRNA 的稳定性

乳腺组织对促乳素（prolactin）的应答提供了一个调控 mRNA 稳定性的好例子。当培养的乳腺组织用促乳素刺激时，以产生乳蛋白质酪蛋白（casein）做出响应。由此可以推测，伴随酪蛋白的增加，酪蛋白 mRNA 浓度也会增加，事实也确实如此。促乳素处理乳腺组织 24h 后，酪蛋白 mRNA 水平增加了约 20 倍。而这并不意味着酪蛋白 mRNA 的合成效率增加了 20 倍。实际上，转录效率只增加了2～3倍。酪蛋白 mRNA 水平的上升主要归因于酪蛋白 mRNA 的稳定性增加了约 20 倍。

Jeffrey Rosen 及同事通过**脉冲示踪**（pulse-chase）实验测定酪蛋白 mRNA 的半衰期。**半衰期**（half-life）是指半数 mRNA 降解所需要的时间。他们在促乳素有或无的条件下，在体内短时间放射性标记了酪蛋白 mRNA，即对细胞进行放射性核苷酸脉冲，使标记的核苷酸掺入转录的 RNA 中。然后将细胞转移到非放射性培养基中，随着标记 RNA 的降解和未标记 RNA 的取代，可跟踪测定 RNA 放射性的丢失。经多次示踪实验后，将酪蛋白 mRNA 同已克隆的酪蛋白基因进行杂交，以测定酪蛋白 mRNA 的水平。标记的酪蛋白 mRNA 消失得越快，表明它的半衰期越短。结果如表 16.1 所示，在促乳素存在条件下，酪蛋白 mRNA 的半衰期迅速增加，从 1.1h 上升到 28.5h。同时，响应促乳素的总 poly(A)$^+$ mRNA 的半衰期仅增加了 1.3～4 倍。看起来促乳素选择性地稳定酪蛋白 mRNA，使酪蛋白的表达极大地增强了。注意，脉冲示踪技术除了用来测定分子半衰期外，还可以做很多事情，如揭示前体-产物关系，因为标记的前体会进入标记的产物中去。我们在本章前面讨论 rRNA 前体和产物的关系时，就列举过一个很好的例子。

表 16.1　促乳素对酪蛋白 mRNA 半衰期的影响

	RNA 半衰期/h	
RNA 种类	－泌乳刺激素	＋泌乳刺激素
rRNA	＞790	＞790
poly(A)$^+$RNA(半衰期短)	3.3	12.8
poly(A)$^+$RNA(半衰期长)	29	39
酪蛋白 mRNA	1.1	28.5

Source: Reprinted from Guyette, W. A., R. J. Matusik, and J. M. Rosen, Prolactinmediated transcriptional and post-transcriptional control of casein gene expression. *Cell* 17: 1013, 1979. Copyright © 1979, with permission from Elsevier Science.

小结　转录后调控基因表达的常见形式是控制 mRNA 的稳定性。例如，当乳腺组织受促乳素刺激时，酪蛋白的合成量会急剧增加，但酪蛋白量的增加主要是由于酪蛋白 mRNA 的半衰期延长了，而不是其基因转录水平的增加。

转铁蛋白受体 mRNA 的稳定性

研究转录后调控最清楚的例子是哺乳动物的铁离子稳态（iron homeostasis）（调控铁离子浓度）。铁离子是所有真核细胞所必需的矿物元素，但高浓度的铁离子对细胞有害。所以，细胞必须精确调节细胞内铁离子的浓度。哺乳动物细胞通过调节两个蛋白质的量来调节铁离子浓度：一个是离子运输蛋白，称为**转铁蛋白受体**（transferrin receptor，**TfR**），另一个是离子储藏蛋白，称为**铁蛋白**（ferritin）。转铁蛋白是离子耐受蛋白，它可通过细胞表面的转铁蛋白受体 TfR 进入细胞。一旦转铁蛋白进入细胞，便将离子传递给细胞内的其他蛋白质，如需要铁离子的细胞色素。另一种情况是，若细胞吸收了过多的铁离子，就以铁蛋白的形式将铁离子储存起来。

因此，当细胞需要较多的铁离子时，细胞会增加转铁蛋白受体的浓度，使更多的铁离子进入细胞，同时减少铁蛋白的浓度，使储存的铁离子不会太多，可利用的更多些。此外，如果一个细胞有太多的铁离子，它会减少转铁蛋白受体的浓度并增加铁蛋白的浓度。细胞采用转录后调控来完成这两个工作，调节铁蛋白 mRNA 的翻译效率，调节转铁蛋白受体 mRNA 的稳定性。我们将在第 17 章讨论铁蛋白 mRNA 的翻译。以下我们先详细讨论转铁蛋白受体 mRNA 的稳定性调节。

1986 年，Joe Harford 及同事报道，利用螯合物耗竭细胞内的铁离子可增加转铁蛋白受体（TfR）的 mRNA 浓度。此外，通过加入血晶素（hemin）或铁盐来增加细胞内铁离子浓度可降低 TfR mRNA 的浓度。细胞内铁离子浓度变化所引起的 TfR mRNA 浓度改变，主要原因不是 TfR mRNA 合成速率的改变，而是在很大程度上依赖于 TfR mRNA 半衰期的改变。具体讲，TfR mRNA 的半衰期可以从铁离子充足时的 45min 增加到离子缺乏时的数小时。我们将查验 mRNA 半衰期数据，但接下来先要搞清楚 mRNA 的结构，因为根据 mRNA 结构才有可能调节其寿命。

铁离子应答元件　1985 年，Lukas Kühn 及同事克隆了一个人类 TfR cDNA，发现其编码 mRNA 有 3 个区域：96nt 的 5′-非翻译区（5′-UTR）、2280nt 的编码区、2.6kb 的 3′-非翻译区（3′-UTR）。为检测较长的 3′-UTR 的作用，Dianne Owen 和 Kühn 删除了 3′-UTR 的 2.3kb，然后用截短的重组子转染小鼠 L 细胞。同时他们还构建了另一个相似的重组子，用 SV40 病毒启动子代替正常的 TfR 启动子。然后用人 TfR 特异性单克隆抗体和荧光标记的第二抗体检测细胞表面的 TfR。图 16.21 显示了检测结果。对于野生型基因，细胞通过增加约 3 倍的表面 TfR 浓度来响应铁螯合剂。他们发现一方面 SV40 启动子调控的 TfR 基因具有相同的表达行为，表明 TfR 启动子与铁离子响应无关；另一方面，删除了 3′-UTR 的基因不响应铁离子，不论铁螯合剂存在与否，细胞表面的 TfR 浓度相同。显然，实验中所删除的 3′-UTR 部分包含铁离子应答元件。

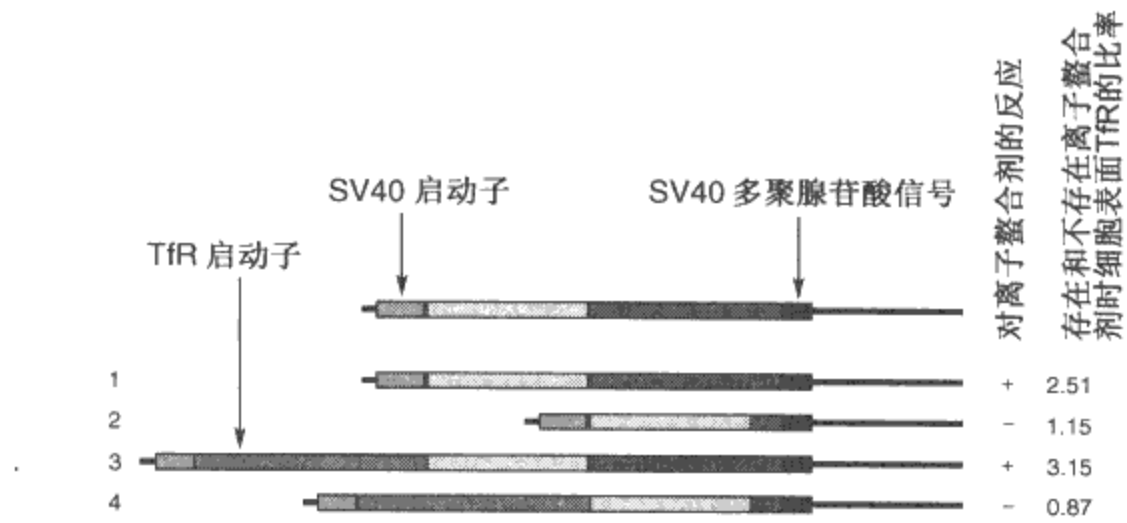


图 16.21　3′-UTR 对 TfR 的细胞表面浓度的铁离子应答的影响。Owen 和 Kühn 构建了如图所示的 TfR 基因。条框中的 DNA 区域用不同颜色表示。SV40 启动子橙色，TfR 启动子蓝色，TfR 5′- UTR 黑色，TfR 编码区黄色，TfR 3′- UTR 绿色，SV40 多聚腺苷酸化信号紫色。用这些质粒分别转染细胞，利用荧光抗体分析细胞表面的 TfR 浓度。右侧数字为铁离子螯合剂（去铁胺）有或无时，细胞表面 TfR 的比率，旁边是细胞对离子螯合剂反应的定性指数（＋或－）。（*Source*：Adapted from Owen，D. and L. C. Kühn，Noncode 3′sequences of the transferrin receptor gene are required for mRNA regulation by iron. *The EMBO Journal* 6：1288，1987.）

当然，细胞表面 TfR 受体的出现并不一定反映 TfR mRNA 的浓度。为了直接检测铁离子对 TfR mRNA 浓度的影响，他们对铁螯合剂处理和未处理细胞的 TfR mRNA 进行了 S1 核酸酶分析（第 5 章）。如预期的那样，铁螯合剂极大地增加了 TfR mRNA 的浓度，但 TfR 基因缺失 3′-UTR 时，对铁离子的这种应答就消失了。

那么 3′-UTR 的哪部分赋予了铁离子应答呢？Harford 及同事发现只需从 3′-UTR 的中部删除 678nt 就可消除大部分的铁离子应答。

计算机分析这段 678nt 片段发现，它最可能包含 5 个发夹或茎环的结构，如图 16.22 所示。更有趣的是，所有这些茎环的总体结构，包括各环的序列，与铁蛋白 mRNA 5′-UTR 中发现的茎环高度相似。该茎环称作**铁应答元件**（iron response element，**IRE**），负责铁离子刺激铁蛋白 mRNA 翻译的能力。言外之意是，这些 IRE 元件是 TfR 基因表达对铁离子应答的中介者。

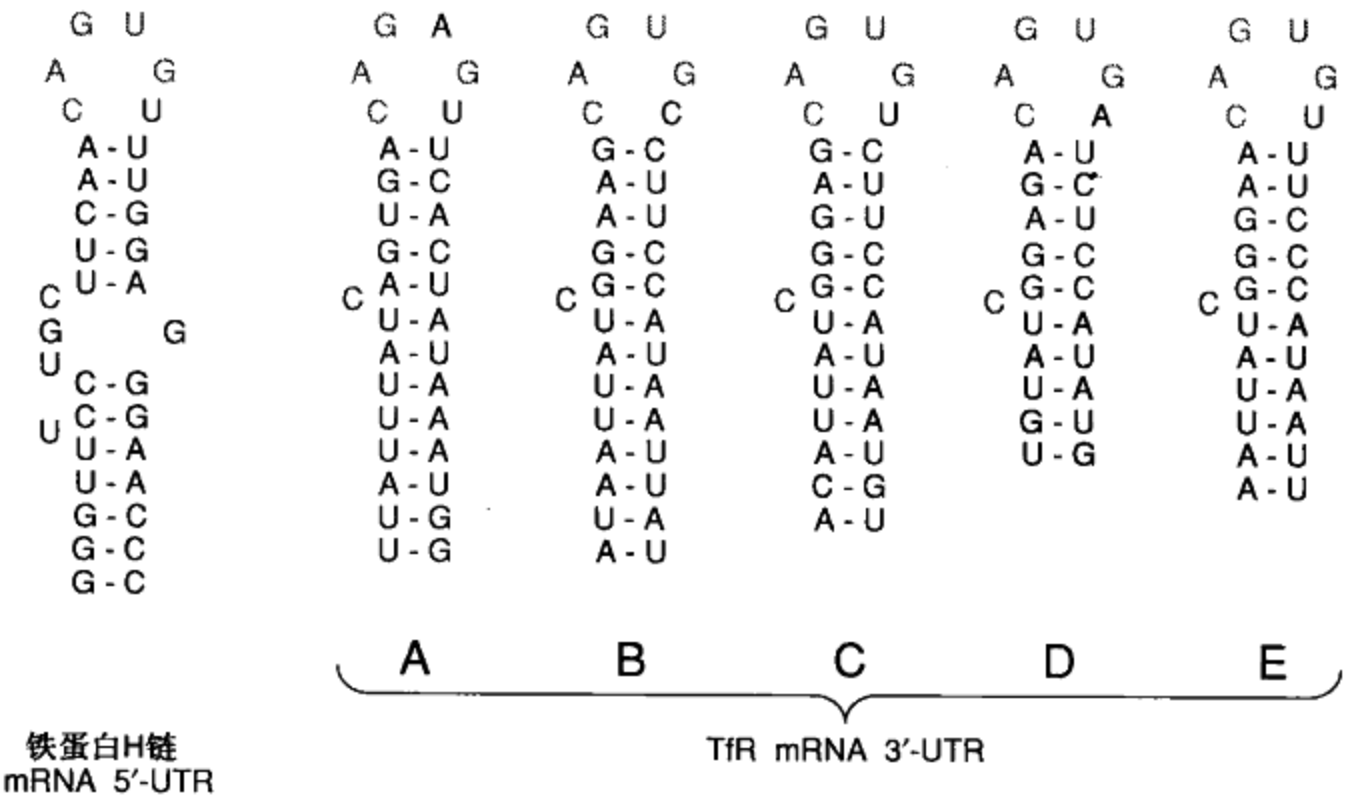


图 16.22　人 TfR mRNA 3′-UTR 的 5 个茎环结构与人铁蛋白 mRNA 5′-UTR 的 IRE 的比较。保守的环突 C 用蓝色表示，环中的保守碱基用红色表示。（*Source*：Adapted from Casey，J. L.，M. W. Hentze，D. M. Koeller，S. W. Caughman. T. A. Rovault，R. D. Klausner，and J. B. Harford，Iron-responsive elements：Regulatory RNA sequences that control mRNA levels and translation. *Science* 240：926，1988.）

Harford 及同事通过凝胶阻滞实验继续研究发现，人细胞中含有一个或多个蛋白质，能特异性地结合人 TfR IRE（图 16.23）。这种结合可被过量 TfR mRNA 或同样具有 IRE 的铁蛋白 mRNA 所竞争，但没有 IRE 的 β-球蛋白 mRNA 不能参与竞争。所以，这种结合是 IRE 特异性的。这一发现强调了铁蛋白和 TfR IRE 存在相似性，提示它们甚至可能结合相同的蛋白质。但正如所看到的，蛋白质与这两个 mRNA 的结合具有不同的作用。

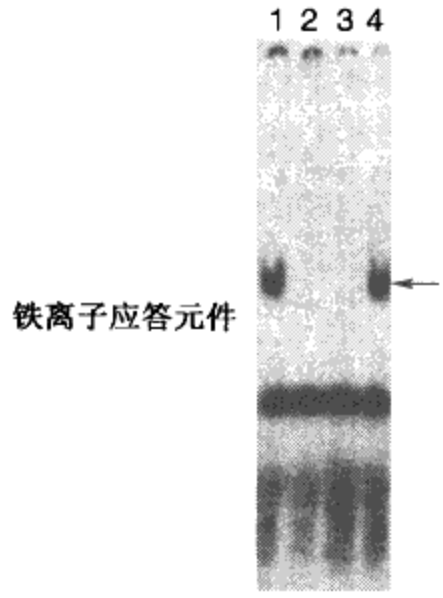


图 16.23　IRE 结合蛋白的凝胶阻滞分析。Harford 及同事制备了标记的 1059bp 转录物，该转录物对应于人的含 5 个 IRE 的 TfR mRNA 的 3′-UTR。将标记 RNA 与人细胞（有或无竞争 RNA）的胞质提取物混合，电泳复合物，放射自显影观察。泳道 1，无竞争物；泳道 2，TfR mRNA 竞争物；泳道 3，铁蛋白 mRNA 竞争物；泳道 4，β-球蛋白 mRNA 竞争物。箭头指示特异性蛋白质-RNA 复合物，可能有一个或多个 IRE 结合蛋白。(*Source*: Koeller, D. M., J. L. Casey, M. W. Hentze, E. M. Gerhardt, L.-N. L. Chan, R. D. Klausner, and J. B. Harford, A. cytosolic protein binds to structural elements within the nonregulatory region of the transferrin receptor mRNA. *Proceedings of the National Academy of Sciences USA* 86 (1989) p. 3576. f. 3.)

小结　铁离子浓度高时转铁蛋白受体 TfR 的浓度就会降低，并且 TfR 浓度的降低主要归因于 TfR mRNA 稳定性的降低。TfR 对铁离子的应答主要依赖于 mRNA 的 3′-UTR，它包含 5 个称为铁离子应答元件（IRE）的茎环结构。

快速周转决定子　已知铁离子通过控制 mRNA 的稳定性来调节 TfR 基因的表达，并且一个蛋白质可以结合到 TfR mRNA 3′-UTR 的一个或多个 IRE 上，所以我们推测 IRE 结合蛋白可以保护 mRNA 免于降解。这种调节要求 TfR mRNA 本身是不稳定的。因为如果它是稳定的 mRNA，通过进一步提高稳定性所得到的相对增益就很少。确实，该 mRNA 不稳定，Harford 及同事证明这种不稳定性是由**快速周转决定子**（rapid turnover determinant）引起的，它也位于 3′-UTR。

什么是快速周转决定子？由于人和鸡 TfR 基因具有相同调控方式，它们可能具有相同的快速周转决定子。所以，比较这两个 mRNA 的 3′-UTR 有可能发现一些共同特征，从而提示我们从哪里开始研究。Harford 及同事比较了人 TfR mRNA 的 678nt 区与鸡 TfR mRNA 的相应区域，发现在包含 IRE 的区域有高度的相似性。图 16.24 的重组子 1 表示人类的 678nt 区结构。两者在该区的 5′端都有两个 IRE，然后是具有大环的茎（人类为 250nt，鸡为 332nt），随后是其他 3 个 IRE。人类 TfR mRNA 中包含 5′-IRE 和 3′-IRE 的区域与鸡 TfR mRNA 的相应区域非常相似，但中间的环区及较远的上游和下游区没有可检测到的相似性。这表明快速周转决定子应该位于 IRE 的某个地方。Harford 等通过突变 TfR mRNA 的 3′-UTR，观察哪些突变体可以稳定 mRNA 鉴定出其中的一些元件。

如图 16.24 所示，他们首先观察了简单的 5′或 3′删除突变体，用这些突变体转染细胞，然后通过比较血晶素或离子螯合剂去铁胺处理后，细胞的 TfR mRNA 和蛋白质水平，以此分析铁离子调控。分别利用 Northern 印迹和免疫共沉淀法检测 mRNA 水平和蛋白质水平的变化。他们发现 250nt 中心环的删除（重组子 2）或 IRE A 的删除（重组子 3）对铁离子调节没有影响。这两个重组子所产生的响应都与野生型（重组子 1）相似，即用铁螯合剂而不是血晶素处理细胞时，TfR mRNA 和蛋白质水平有很大提高。但是，同时删除 IRE A 和 B 后（重组子 4）消除了铁离子调节作用，在两种处理条件下 TrR mRNA 和蛋白质水平不变（都很高）。因此，缺失 IRE B 时 TrR mRNA 是稳定的，所以该 IRE 可能是快速反转决定子的一部分。3′删除得到相似的结果。删除 IRE E 对铁调节几乎没有影响，而同时删除 IRE D 和 E 可稳定 TfR mRNA，甚至在血晶素存在情况下也如此。因此，IRE D 可能是快速周转决定子的一部分。

基于以上结果，可以预测 IRE A、IRE E 和中心环的删除不改变铁离子的调控。于是，他们人工合成了一个缺失这 3 个部分的元件，叫做 TRS-1，如图 16.25a 所示。正如预期的那样，包含这些元件的 mRNA 完全保留了铁离子应答功能。接着，他们制备了两个 TRS-1 的突变体（图 16.25a）。一个是

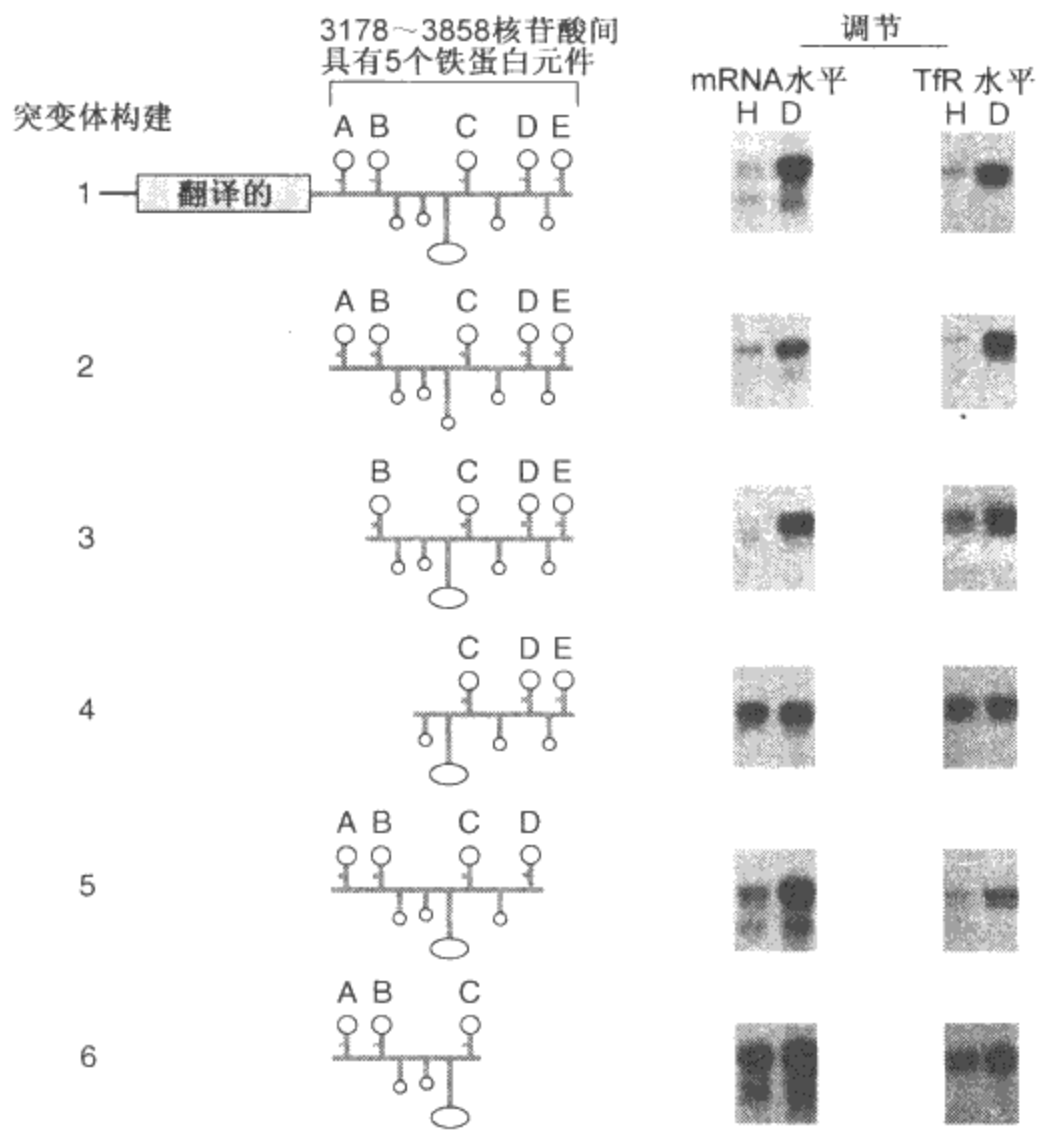


图 16.24　用 3′-UTR 删除突变体分析铁离子应答性。Harford 及同事在人 TfR 基因 3′-UTR 的部分区域做删除突变，该区含 IRE 长约 678nt。用这些重组子转染细胞（1～6）。用血晶素（H）或去铁胺（D）处理后，用 Northern 印迹测定 mRNA 浓度（mRNA 水平）。同样处理后，用免疫共沉淀测定 TfR 蛋白浓度（水平）。Northern 印迹与免疫沉淀结果如右图所示。(*Source*: Casey, J. L., D. M. Koeller, V. C. Ramin, R. D. Klausner, and J. B. Harford, Iron regulation of transferrin receptor. *EMBO Journal* 8 (8 Jul 1989) p. 3694. f. 1.)

TRS-3，3 个 IRE 都删除了，剩下的是其余的茎环，即图 16.25 中向下画出的茎环。另一个是 TRS-4，仅丢失了 3 个碱基，它们是分别位于每个 IRE 环 5′ 端的 C。图 16.25b 显示了这两个突变体的效果。无 IRE 的 TRS-3 失去了全部铁离子应答，并且 TfR RNA 看起来比野生型 mRNA 稳定得多。换句话说，即使有血晶素存在，仍有大量的 TfR。TRS-4 是每个 IRE 都缺失一个 C，丧失了大部分铁离子应答，但 mRNA 保持不稳定。换句话说，即使在铁螯合剂存在时仍然没有多少 TfR。因此，该 mRNA 保留了其快速周转决定元件，但失去了被 IRE 结合蛋白稳定的能力。事实上，如我们所期望的那样，凝胶阻滞实验显示 TRS-4 不能结合 IRE 结合蛋白。

为了进一步确定快速周转决定子，Harford 及同事构建了两个新的重组子，他们分别删除了大中心茎环两边的（方向朝下的）非 IRE 茎环。通过细胞转染和免疫共沉淀实验（如前面所述）检测重组子。两个重组子几乎都丢失了所有的铁应答，都有组成型 TfR 高水平的表达（与图 16.25 中 TRS-3 的影响相同）。因此，两个茎环的缺失对赋予 mRNA 快速周转的能力都具有重要作用。为了证明该结果不是由 mRNA 与 IRE 结合蛋白相互作用的不稳定性造成的，研究人员利用凝胶阻滞实验（如前面所述）检测了蛋白质-RNA 的结合。两个突变体都能够像野生型 mRNA 那样同 IRE-结合蛋白进行结合，并且这种结合在任何情况下都受到非标记的 IRE 竞争。

小结　缺失 IRE A 和 E 以及 TfR 3′-UTR 中间的大茎环不改变铁应答。然而，缺失所有的 IRE，或两个非 IRE 茎环中的任何一个都会使 TfR mRNA 变得稳定。所以，非 IRE 茎环中的任何一个，和至少 IRE B～D 中的一个，都是快速周转决定子的一部分。从 IRE B～D 移除一个胞嘧啶核苷酸导致 TfR mRNA 变得不稳定且不能结合 IRE-结合蛋白。

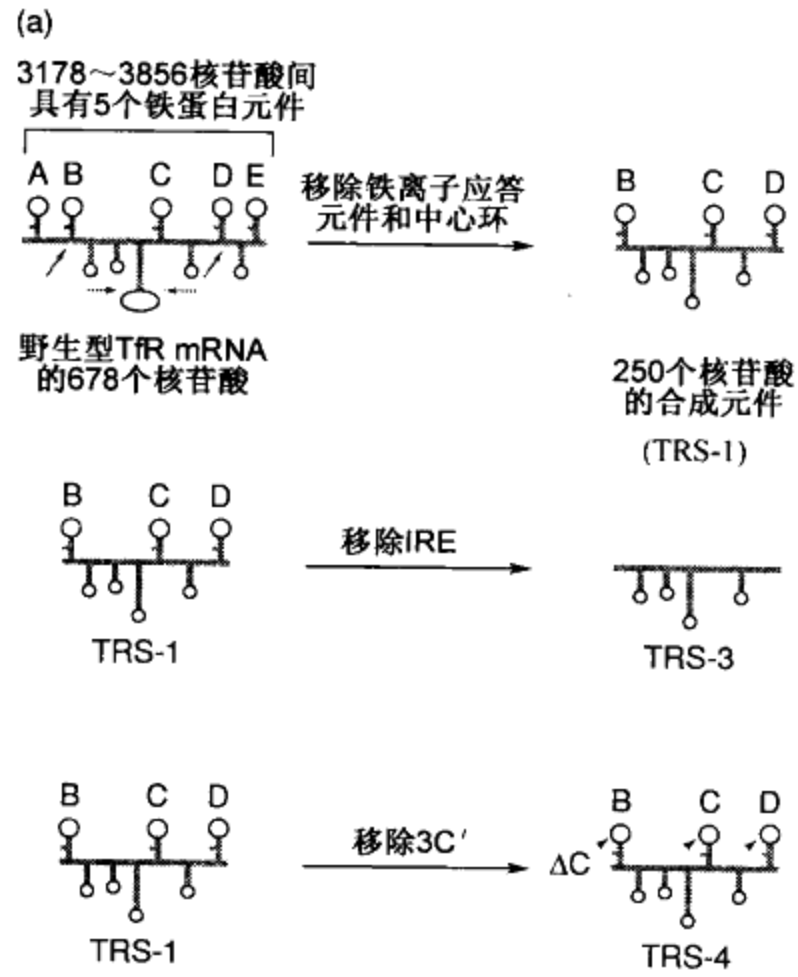


图 16.25　TfR 3′-UTR 的 IRE 区的删除对铁离子应答的影响。（a）删除突变体的产生。Harford 及同事切除 IRE 的 A 和 E 及大中间环得到 TRS-1 突变体，如图中箭头所示。切除 TRS-1 余下的 3 个 IRE 得到 TRS-3 突变体，切除每个 IRE 环 5′端的一个 C 得到 TRS-4 突变体。（b）检测突变体的铁离子响应。分别用重组子转染细胞，然后，一半细胞用血晶素（H）处理，另一半用去铁胺（D）处理，免疫沉淀法分析 TfR 的生物合成。放射自显影结果如图所示，转染的突变体和铁离子处理如图上方所示。铁调节百分比在图下方给出。用各处理的野生型对照，以 100%表示。结果表明，铁螯合剂与血晶素的处理（D/H）是成倍诱导。TRS-3 没有调节作用，而且 TfR 的组成型合成水平很高，说明其 mRNA 稳定。TRS-4 调节作用有限，TfR 合成水平低，说明其 mRNA 不稳定。(*Source*: Casey, J. L. , D. M. Koeller, V. C. Ramin, R. D. Klausner, and J. B. Harford, Iron regulation of transferrin receptor mRNA levels requires iron-responsive elements and a rapid turnover determinant in the 3′ untranslated region of the mRNA. *EMBO Journal* 8(8 Jul 1989)p. 3695, f. 3B.)

TfR mRNA 稳定性的决定　到目前为止，我们已经知道 TfR mRNA 的水平受细胞内铁浓度的调节，并且这一应答由 3′-UTR 决定，而不是启动子。这也清楚地表明铁离子调节了 TfR mRNA 的半衰期，而不是 mRNA 的合成速率，但是我们还没有提供明确的证据来证明这一假说的正确性。为了提供这种证据，Ernst Mullner 和 Lukas Kuhn 在铁螯合剂去铁胺有或无的情况下，检测了 TfR mRNA 的降解速率。发现铁浓度低时，TfR mRNA 非常稳定。而在高铁浓度时，TfR mRNA 的降解非常快。这两种情况下，TfR mRNA 的半衰期分别是 30h 和 1.5h，所以铁离子似乎使 TfR mRNA 的稳定性下降为原先的大约 1/20。

Harford 和他的同事通过检测重组子 TRS-A、TRS-3 和 TRS-4（图 16.25）得到了相似的结果。他们用这些重组子转染细胞，然后用血晶素或铁螯合剂处理细胞，检测其 TfR mRNA 的半衰期，步骤如下：用放线菌素 D（actinomycin D）阻断 mRNA 的进一步合成，然后在处理后的不同时间分离细胞的 RNA，Northern 印迹检测 TfR mRNA 浓度。对 TRS-1mRNA（具有完全铁应答活性），得到如下结果：有血晶素时，mRNA 的半衰期大约是 45min，而铁螯合剂去铁胺大大提高了 TfR mRNA 的稳定性。在这两种条件下，mRNA 的水平在药物处理后的 180min 还没有明显的下降。当铁离子移走后 mRNA 的半衰期长于 3h。

Harford 和合作者检测了经血晶素或者铁螯合剂处理突变体 TRS-A、TRS-3、TRS- 4 和 TRS-1 的 mRNA 半衰期。如前面所述，当铁浓度提高时，TRS-1 mRNA 不稳定，但是在铁离子被螯合剂去除后却是稳定的。失去了 IRE 和快速周转决定子的 TRS-3 在所有条件下都是稳定的。相反，TRS-4 具有突变的 IRE，它不能结合 IRE-结合蛋白，但它的快速周转决定子仍是完整的，所以不管细胞内铁离子浓度如何它都是不稳定的。

小结　当铁离子浓度升高时，TfR mRNA 快速降解。当铁离子浓度降低时，TfR mRNA 降解速率减慢。mRNA 的稳定性相差约 20 倍，mRNA 稳定性的差异比不同铁浓度时 mRNA 合成的差异（约 2 倍）在决定 TfR mRNA 的积累速率上更重要。没有快速周转决定子的 TfR-3 mRNA 始终是稳定的（在低和高铁浓度下）。不能结合 IRE 结合蛋白的 TfR-4 mRNA 始终是不稳定的。

TfR mRNA 降解途径　Harford 和他的同事研究了 TfR mRNA 的降解机制，发现降解过程首先是在 IRE 区域产生一个内切核酸切口。与其他很多 mRNA 降解途径不同的是，在 TfR 降解开始之前似

乎不需要去腺嘌呤化［移除 poly（A）］。

研究人员首先使用血晶素处理人类浆细胞（ARH-77），然后利用 Northern 印迹法检测 TfR mRNA 的降解情况，发现 TfR mRNA 的水平在处理 8h 内急剧下降。将放射自显影曝光的时间延长后，他们又检测到一个新的 RNA 类型，1000～1500nt，比全长 TfR mRNA 短一些，是在 TfR mRNA 降解过程中出现的（图 16.26a）。这个 RNA 也存在于去poly(A)的片段中（图 16.26b），表明它已失去其 poly（A）尾。但是这个变短的 RNA 的长度分析表明它并不只是丢失了 poly（A）。最简单的解释是一个内切核酸酶在其 3′-UTR 进行切割，去除了超过 1000nt 的 3′端核苷酸，其中包括 poly（A）。

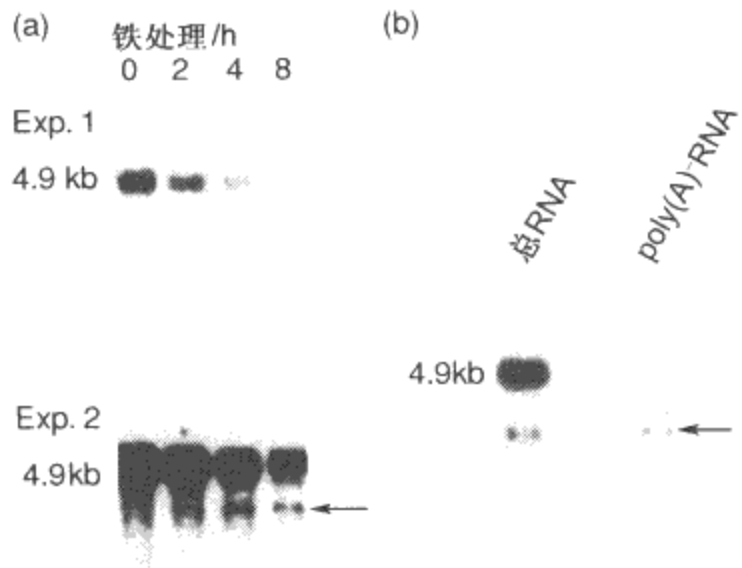


图 16.26　血晶素处理后出现变短的 TfR mRNA。（a）人细胞中缩短转录物的检测。Harford 及同事分离了血晶素处理不同时间的 ARH-77 细胞的 RNA，并进行电泳和 Northern 印迹分析，TfR 探针（Exp1）杂交，短时间放射自显影曝光可检测到长约 4.9kb 的全长转录物（Exp2）。长时间曝光，除了可以检测到长的转录物，还可以检测到一个短的 RNA 产物。（b）在 poly（A）$^-$ 片段中有更短的 RNA。与（a）中的方法一样，RNA 在血晶素中曝光 6h 后，进行 oligo（dT）纤维素色谱分析，以分离 poly（A）$^-$ 与 poly（A）$^+$ RNA。然后对总 RNA 与 poly（A）$^-$ RNA 进行与（a）相同的分析。poly（A）$^-$ 片段只含有缩短的 RNA 片段。（*Source*：Binder，R.，J. A. Horowitz，J. P. Basilion，D. M. Koeller，R. D. Klausner，and J. B. Harford，Evidence that the pathway of transferrin receptor mRNA degradation involves an endonucleolytic cleavage within the 3′ UTR and does not involve poly（A）tail shortening. *EMBO Journal* 13（1994）p. 1970，f. 1.）

为了检测这一假说，Harford 小组用去铁胺处理含有 TRS-1 或 TRS-4 突变体的细胞以除去铁离子，然后提取 RNA，用寡聚核苷酸和 RNase H 在翻译终止密码子附近切割 TfR mRNA（如第 14 章所述）。然后用标记引物进行引物延伸实验，该引物与 SV40 序列杂交，而 SV40 序列超出了 TfR 编码区域的 3′-UTR 末端。所以，这个引物扩增分析可以检测到 mRNA 的 3′部分，而 Northern 杂交可以检测其 5′部分。RNase H 酶切割的目的是缩短 RNA 的长度，产生一个足够小的引物延伸产物，以便可以通过凝胶电泳进行检测。

引物延伸实验显示，TRS-1 和 TRS-4 都有一个清晰的全长 PCR 信号，以及每个 RNA 的较短片段信号。TRS-4 产物的短片段的信号比 TRS-1 产物的更明显。如果该片段代表了 TfR mRNA 的降解产物，那么结果正是我们所期望的，因为有铁螯合剂时，TfR-4 RNA 比 TfR-1 RNA 更不稳定。这个较短的引物延伸产物的大小正好将其 5′端定位在 IRE C 和 D 之间，这与 TfR mRNA 降解的假说一致，即发生降解的第一步是在内切核酸酶作用下从 mRNA 的 3′端产生一个超过 1000nt 的片段。

我们所有的实验结果都与该假说一致（图 16.27）：当铁浓度低时，IRE 结合蛋白或**铁调节蛋白**（iron regulatory protein，**IRP**）在 TfR mRNA 的 3′-UTR 与快速周转决定子结合，保护 mRNA 免受降解。当铁浓度高时，铁结合到 IRE 结合蛋白上，使其从快速周转决定子上解离，使它敞开以便让特异性内切核酸酶切割，该酶从 TfR mRNA 的 3′端切去约 1kb 的片段，使 mRNA 稳定性降低，迅速降解。

顺乌头酸酶（aconitase）是一种既可以结合转铁蛋白受体 mRNA 的 IRE，又可以结合铁蛋白 mRNA 中 IRE 的蛋白质（IRP1），它在柠檬酸循环中可将柠檬酸转化为异柠檬酸。顺乌头酸酶的活化形式是含铁蛋白质，不能结合到 IRE 上。但是，缺乏铁的顺乌头酸酶的脱辅基形式（apoprotein）可以与 mRNA 的 IRE 结合。

小结　TfR mRNA 降解的起始事件可能是在 mRNA 中距 3′端 1000nt 的位置对 RNA 进行的内切核酸酶切割，且其切割位点在 IRE 区域内。这种切割不需要 mRNA 提前去腺嘌呤化。铁控制 TfR mRNA 稳定性的机制如下：当铁浓度低时，至少部分顺乌头酸酶以不含铁的脱辅基形式存在。该蛋白质结合到 mRNA 的 IRE 上，保护 RNA 免受 RNase 攻击。当铁浓度高时，顺乌头酸脱辅基酶结合铁后就不能结合 TfR mRNA IRE，使 mRNA 极易被降解。

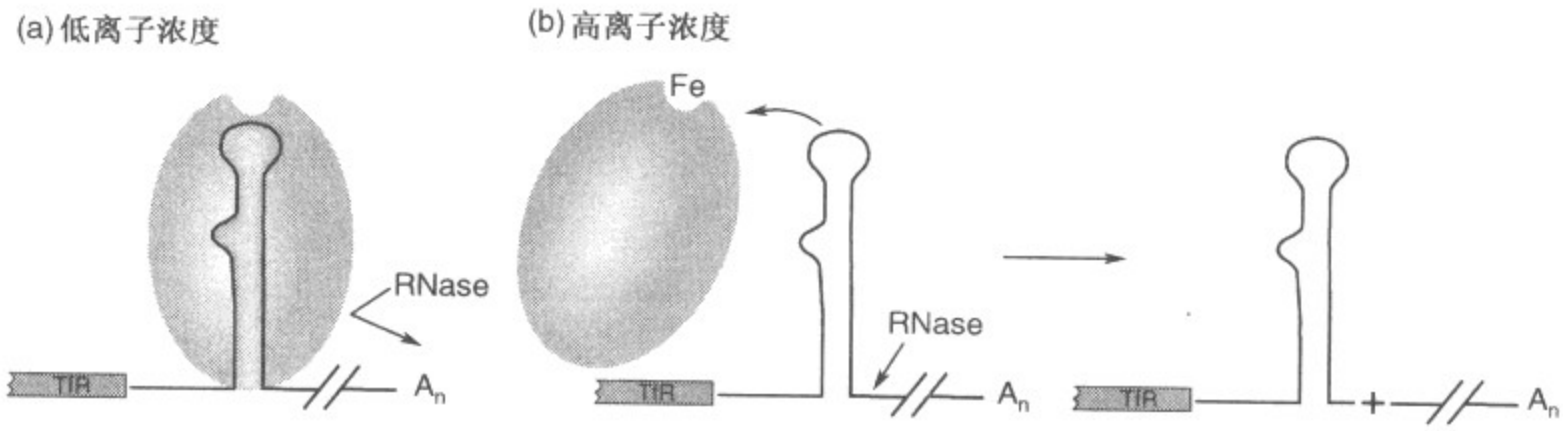


图 16.27　铁介导 TfR mRNA 不稳定性的模型。（a）在低铁浓度下，顺乌头酸酶脱辅基蛋白（橙色）在 TfR mRNA 的 3′-UTR 区与 IRE 结合，阻断 RNase 降解 TfR mRNA。(b) 在高铁浓度下，铁与顺乌头酸酶脱辅基蛋白结合，使之离开 IRE，RNase 便可攻击敞开的 IRE。RNase 至少切割 mRNA 一次，暴露出 mRNA 的 3′端，以便进一步降解。

RNA 干涉

很多年以来，分子生物学家就开始利用反义 RNA 抑制活细胞内特定基因的表达。基本原理是与 mRNA 互补的反义 RNA，可与 mRNA 碱基配对进而抑制其翻译。这种策略通常可以发挥作用，但这种推理又不完整。Su Guo 和 Kenneth Kenphues 在 1995 年将正义 RNA 链注射到细胞后，它阻断特殊基因表达的效果与反义 RNA 一样好。随后，Andrew Fire 及同事在 1998 年发现，双链 RNA（dsRNA）比正义或反义 RNA 具有更好的抑制效果。事实上，正义和反义 RNA 能发挥作用的主要原因是存在少量污染（或产生的）dsRNA，实际上是这种 dsRNA 在抑制基因表达中发挥了最重要作用。

此外，从 1990 年开始，分子生物学家就注意到在不同的转基因生物体内，有时会表现与期望相反的效应。有时生物体不但没有开启反而关闭了转入的基因，甚至是该基因正常的细胞拷贝。其中一个早期例子是研究人员试图通过增加色素形成基因在细胞中的拷贝数来增强矮牵牛花的紫色。但是 25% 转基因植物的花是白色的或紫色和白色相间的，这与期待的效果刚好相反（图 16.28）。对这种现象曾有多种名称，在植物中称为**共抑制**（cosuppression）和**转录后基因沉默**（posttranscriptional gene silencing，**PTGS**），在动物如线虫（秀丽隐杆线虫）和果蝇中称为 **RNA 干涉**（RNA interference，**RNAi**），在真菌中则称为**猝灭**（quelling）。为了避免概念的混淆，从现在起不论研究什么物种，都将这种现象称为 RNAi。

RNAi 的机制　Fire 及同事发现，将 dsRNA（**触发 dsRNA**，trigger dsRNA）注射到线虫生殖腺中，会在胚胎中发生 RNAi，并且在发生 RNAi 的胚胎中检测出相应 mRNA（**靶 mRNA**）的丢失（图 16.29）。但是，dsRNA 必须有外显子区，与内含子

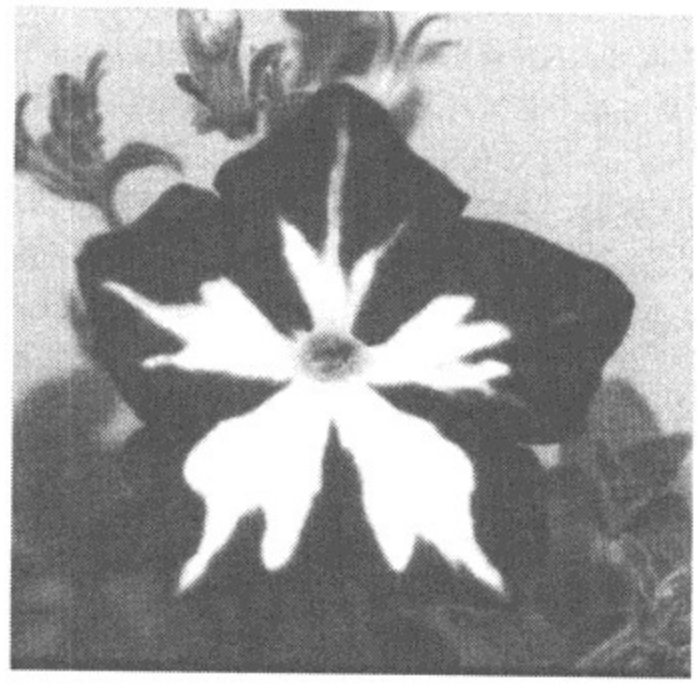


图 16.28　矮牵牛中额外增加颜色基因的拷贝数导致紫色基因表达沉默。花瓣白色条纹的位置发生了基因沉默。(*Source*: Courtesy of Dr. Richard A. Jorgensen, The Plant Cell.)

和启动子序列对应的 dsRNA 不会产生 RNAi。最后，他们证明 dsRNA 的作用可以跨细胞界限，至少在线虫中，说明这种效应普遍存在于整个生物界。

那么，对相应 dsRNA 做出响应的特异性 mRNA 丢失，是由于基因转录抑制还是由于 mRNA 的降解呢？1998 年，Fire 及同事以及其他人证明，RNAi 是一个涉及 mRNA 降解的转录后加工。一些研究者报道，在经历 RNAi 的细胞中存在小片段的 dsRNA，被称为**短干涉 RNA**（short interfering RNA，**siRNA**）。2000 年，Scott Hammond 和合作者从经历 RNAi 的果蝇胚胎中纯化了一种核酸酶，可消化目标 mRNA。具有核酸酶活性的部分纯化的制备物中也含有 25nt 的 RNA 片段，该片段可以被 Northern 印迹检测到，所用探针为该核酸酶靶 mRNA的有义链或无义链。用微球菌核酸酶降解 25nt

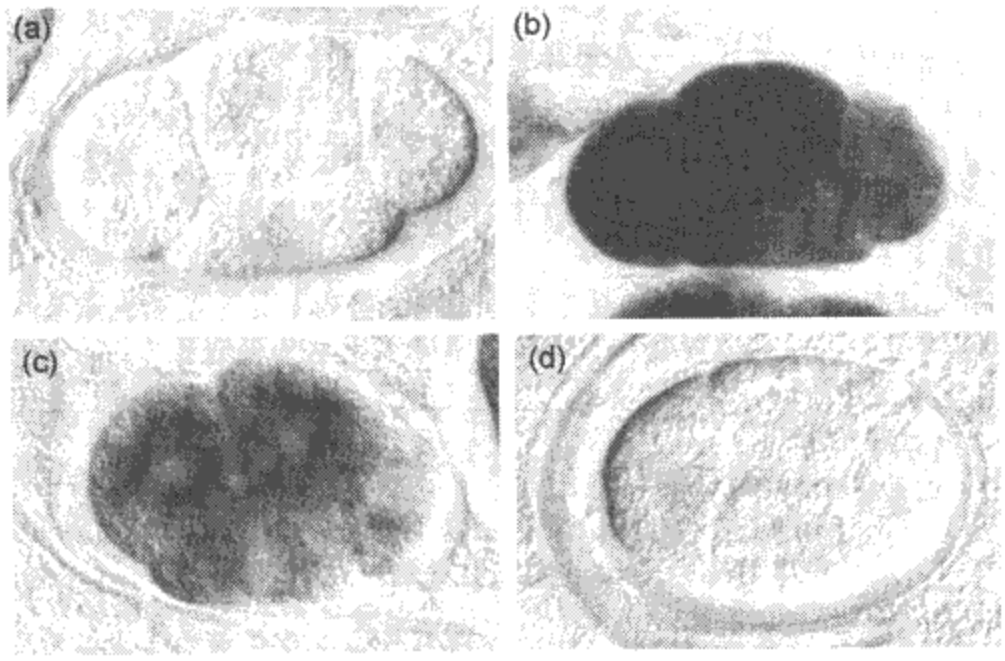


图 16.29　dsRNA 诱导的 RNAi 导致特异性 mRNA 的降解。Fire 及同事将对应线虫 *mex-3* mRNA 的反义 RNA 或 dsRNA 注射到线虫卵巢中。24h 后，将胚胎移植到处理过的卵巢中，用 *mex-3* mRNA 探针进行原位杂交（第 5 章）。(a) 无杂交探针的阴性对照的亲本胚胎。(b) 无注射 RNA 的阳性对照的亲本胚胎。(c) 注射了 *mex-3* 反义 RNA 的胚胎，有大量 *mex-3* mRNA 残留。(d) 注射了与 *mex-3* mRNA 部分对应的 dsRNA 的亲本胚胎，未检测到 *mex-3* mRNA 的残留。(*Source*: Fire, A., S. Xu. M. K. Montgomery, S. A. Kostas, S. E. Driver, and C. C. Mello, Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391 (1998) f. 3, p. 809. Copyright © Macmillan Magazines Ltd.)

的 RNA 能摧毁制备物降解 mRNA 的能力。这些结果表明，核酸酶将起始 RNAi 的 dsRNA 降解成 25nt 的片段，然后这些片段与核酸酶结合并提供引导序列，使核酸酶结合到相应的靶 mRNA 上。

Phillip Zamore 等根据果蝇胚的裂解产物建立了一个可在体外进行 RNAi 的体系。通过该体系可观察到 RNAi 过程的每个步骤。所用胚胎注射了对应萤光素酶（luciferase）mRNA 的触发 dsRNA，研究者可以将萤光素酶 mRNA 靶向降解。首先，Zamore 及其合作者证明 RNAi 过程需要 ATP。他们将己糖激酶（hexokinase）和葡萄糖与提取物共孵育，耗竭其中的 ATP，因为己糖激酶可将 ATP 转化为 ADP，并将脱离下来的磷酸基转移给葡萄糖。ATP 耗竭的提取物不再继续降解目标萤光素酶 mRNA 了。

接着，他们标记了 dsRNA 的一条链（或两条链），结果发现无论标记哪条链，都会出现标记的 21～23nt siRNA（图 16.30）。siRNA 的产生不需要 mRNA（比较泳道 2 和 3），所以这些短 RNA 显然来自 dsRNA，而不是 mRNA。当加帽的反义萤光素酶 RNA 被标记时（泳道 11 和 12），少量 siRNA 出现，且其数量在 mRNA 存在时会有所增加（泳道 12）。这一结果提示，标记的反义 RNA 与加入的 mRNA 杂交产生了 dsRNA，dsRNA 可被降解为短 RNA 片段。总之，所有这些结果表明，存在一个核酸酶将触发 dsRNA 降解成短片段。进一步的研究还表明这些 siRNA 长 21～23nt。

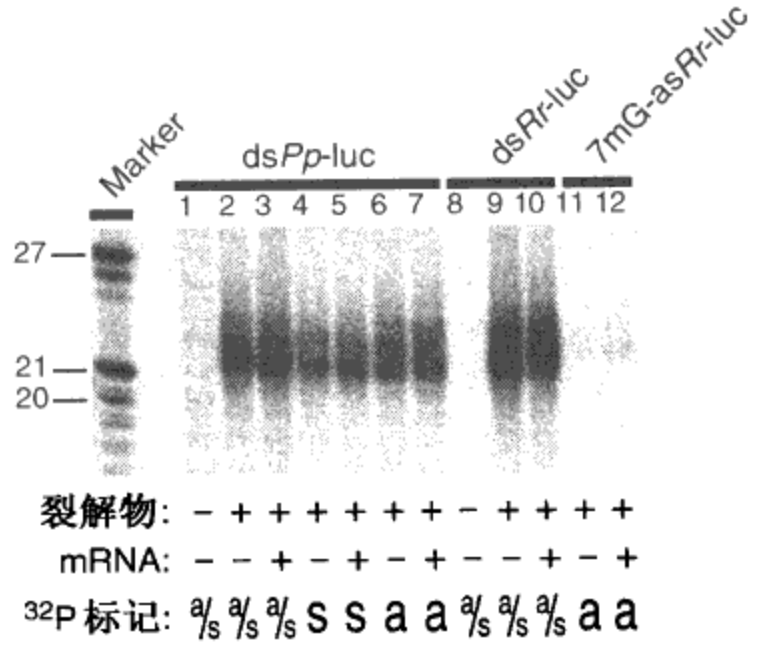


图 16.30　在 RNAi 感受态果蝇胚胎提取物中21～23nt RNA 片段的生产。如图顶部所示，Zamore 及其合作者将萤火虫（*Photinus pyralis*）（*Pp*-luc RNA）或水母（*Renilla reniformis*）（*Rr*-luc RNA）的双链萤光素酶 RNA 加入有或无相应 mRNA 的裂解物中，如图底部所示。分别用标记的有义链（s）、无义链上（a）或两条链（a/s）都标记的 dsRNA。17～27nt 的 RNA 标准在左边泳道（Marker）给出。泳道 11 和 12 分别是有和无 mRNA 时标记的、加帽的反义 *Rr*-luc RNA。(*Source*: Zamore, P. D., T. Tuschl, P. A. Sharp, and D. P. Bartel, RNAi: Double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* 101 (2000) f. 3, p. 28. Reprinted by permission of Elsevier Science.)

然后，Zamore 和他的同事证明，触发 dsRNA 决定靶 mRNA 的切割位置。他们向 RNAi 提取物中加入 3 个不同的触发 dsRNA，其末端相差 100nt，然后加入 5′-标记的 mRNA 进行 RNA 切割，电泳产物。图 16.31 显示电泳结果。dsRNA（c），它的 5′端距 mRNA 的 5′端最近，所产生的片段最短。dsRNA（B）其 5′端位于 dsRNA 5′端下游约 100nt 处，它产生一个比 dsRNA（C）的片段约长 100nt 的 mRNA片段。dsRNA(A)，它的 5′端位于 dsRNA（B）5′端下游约 100nt 处，产生一个比 dsRNA（B）的片段约长 100nt 的 mRNA 片段。触发 dsRNA 相对于 mRNA 的位置和切割位点之间的紧密关系，有力地证明了 dsRNA 决定 mRNA 的切割位点。

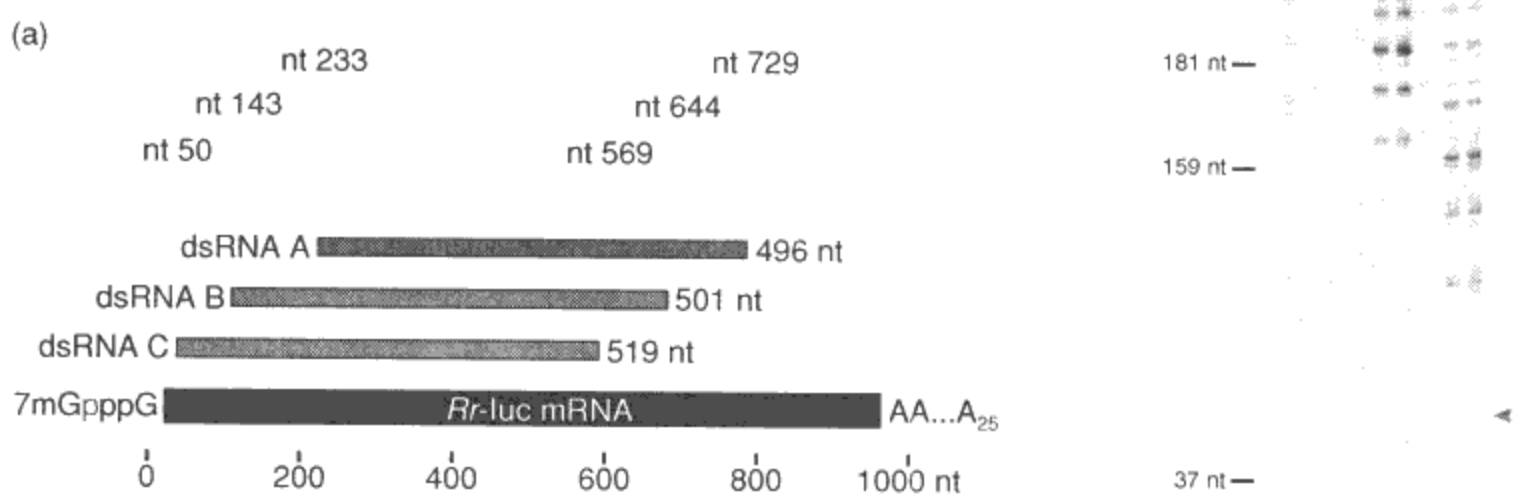


图 16.31　触发 RNA 引导 mRNA 的剪切。Zamore 及同事将图（a）中 3 种 dsRNA 加到含有 *Rr*-luc mRNA 的胚胎提取物中，该 mRNA 的 5′帽子中含有标记的磷酸基团。(b) 实验结果。电泳 5′端标记 mRNA 的降解产物。实验中产生的 dsRNA 在图顶部标出，并用颜色区分。第一泳道，标号为 0，不含 dsRNA。反应时间（以 h 表示）如图顶端所示。箭头表示 RNA C 位置之外的一个弱剪切位点。其余的剪切位点都在 mRNA 上 3 个 dsRNA 所对应的区域内。(*Source*：Zamore，P. D.，T. Tuschl，P. A. Sharp，and D. P. Bartel，2000. RNAi：Double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* 101（2000）f. 5，p. 30. Reprinted by permission of Elsevier Science.)

接下来，Zamore 及同事对来自图 16.31 的 mRNA 降解产物进行了高分辨率凝胶电泳。结果非常惊人（图 16.32）。在 mRNA 中，主要的切割位点大部分以 21～23nt 为间隔，产生一系列 mRNA 片段，其长度的差异是 21～23nt 的倍数。唯一明显的例外在箭头所指位点，它距前一个切割位点仅有 9nt。该特殊位点处于 7 个尿嘧啶的碱基串中。这个发现很有趣，因为 16 个切割位点中的 14 个也被定位在尿嘧啶处。除了这个特例外，21～23nt 的间隔恢复了剩下的切割位点图谱。这些结果支持了21～23nt 的 siRNA 决定 mRNA 在何处切割的假说，并且表明切割位点倾向于在尿嘧啶处。

2001 年，Hammond 及同事报道，他们从果蝇中分离纯化到将触发 dsRNA 切割为短片段的酶，并称之为 **Dicer**，因为该酶将双链 RNA 切成大小均一的片段。Dicer 是本章前面所提到的 RNase III 家族的成员。实际上，Hammond 及同事是直接在 RNase III 家族寻找 Dicer 的，因为 RNase III 是目前已知唯一对 dsRNA 特异性切割的核酸酶。像 RNase III 那样，Dicer 在双链 siRNA 的末端产生 2nt 的 3′黏端（突出的 3′端）以及磷酸化的 5′端。

3 个早期证据已表明 Dicer 在 RNAi 中对 RNA 进行切割。首先，编码 Dicer 的基因 *Dicer* 表达产生可将 dsRNA 切成 22nt 小片段的蛋白质。其次，由该蛋白质制备的抗体可与果蝇提取物中的一个酶结合，该酶将 dsRNA 切割成小片段。最后，将 *dicer* dsRNA 导入果蝇细胞中，能部分阻断 RNAi。具有讽刺性的是，Hammond 及同事可以用 RNAi 阻断

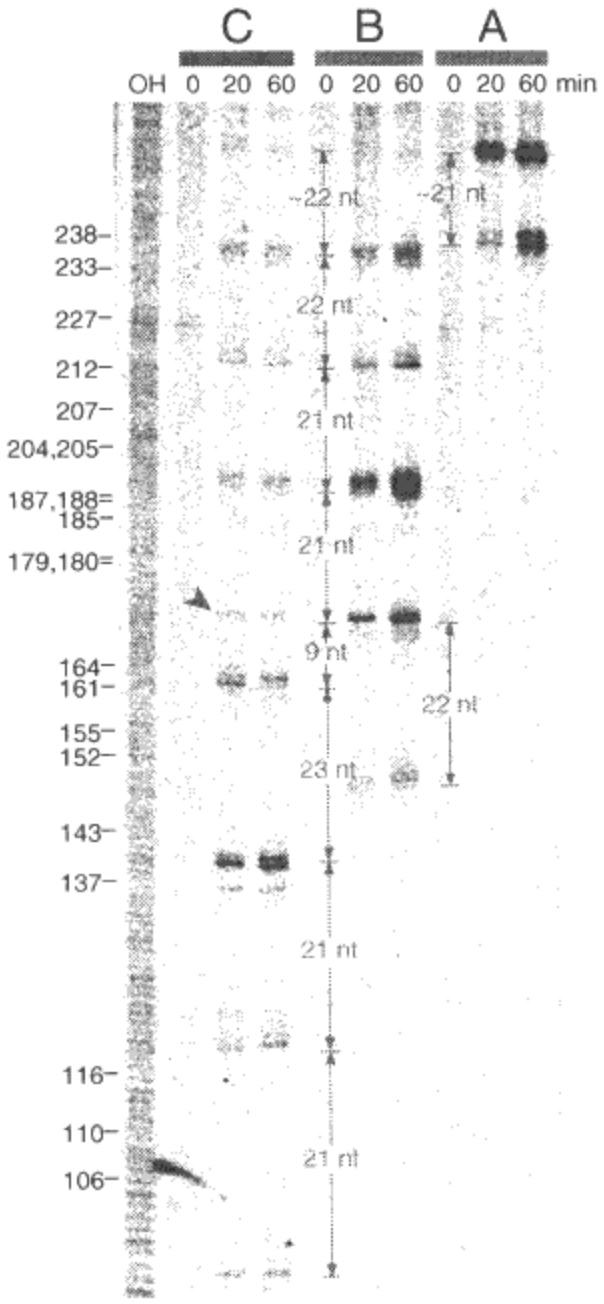


图 16.32　RNAi 中靶 mRNA 的剪切位点间隔为 21～23nt。Zamore 及同事将含有图 16.31 中 3 种触发 dsRNA 的 RNAi 产物进行了高分辨率变性聚丙烯酰胺凝胶电泳分析。剪切位点间隔 21～23nt，除箭头指明的特例以外。这个特异带为 9nt 间隔，在这以后的剪切间隔又恢复了 21～23nt。（*Source*: Zamore, P. D., T. Tuschl, P. A. Sharp, and D. P. Bartel, RNAi: Double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* 101 (2002) f. 6, p. 31. Reprinted by permission of Elsevier Science.）

RNAi。但是如果你仔细分析就会发现这种阻断永远不会完成。

Dicer 还具有 RNA 解旋酶活性，可以将由它产生的 siRNA 的两条链分开，至少在理论上如此。然而，Dicer 不能行使 RNAi 的第二步，即不能切割靶 mRNA。这是被称为 **slicer** 的另一个酶的工作，slicer 存在于 **RNA 诱导的沉默复合物**（RNA-induced silencing complex，**RISC**）中。图 16.33 总结了目前我们所了解的 RNAi 的机制。

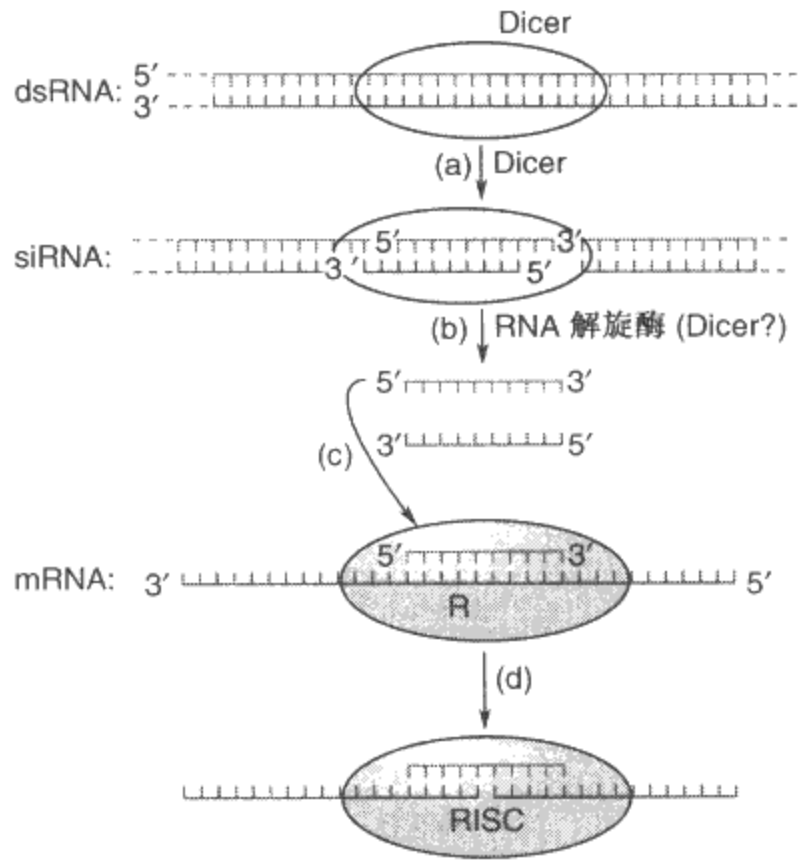


图 16.33　RNAi 简要模型。(a) Dicer（黄色）识别并结合 dsRNA（红色和蓝色），将 dsRNA 剪切成 21～23nt 的 siRNA（简便起见，此处仅用 10nt 表示），带有 2nt 的 3′端突出。中间 siRNA 的末端被标记以便显示 3′端突出。(b) Dicer 的解旋酶活性将 dsRNA 分离为单链。(c) siRNA 中的一条链（红色）同 RISC 复合体（橙色）结合，通过碱基配对与目标 mRNA 结合。(d) RISC 复合体中的 siRNA 单链作为引导链指导目标 mRNA的切割，切割位点在 siRNA 互补链的中间。

Hammond 和其他人通过遗传学实验表明，另一个果蝇蛋白 **Argonaute** 在 RNAi 的第二步（slicer）反应中是必需的。Argonaute 不具有 RNase III 基序，所以分子生物学家起初排除其作为 slicer 的候选蛋白。但是，Leemor Joshua-Tor 和 Gregory Hannon 及同事在 2004 年对 Argonaute 的结构、生化性质和遗传功能的研究证明，它基本上具有 slicer 活性。

他们在 2003 年的结构研究发现，果蝇 Argonaute2 具有两个特征性结构域，**PAZ** 和 **PIWI**（PAZ 取名于 PIWI、Argonaute 和 Zwili，PAZ 只在 Argonaute 和 Dicer 中发现；PIWI 取名于蛋白 piwi）。PAZ 含有与所谓的 OB 折叠相似的模块，可结合单链 RNA。他们还证明 PAZ 结构域可结合单链的 siRNA，或双链 siRNA 3′端的 2nt 单链突出端，所用方法是标记的 siRNA 和克隆表达的 GST-PAZ 融合蛋白进行交联实验。这表明在 slicer 反应中，Argonaute 至少可以作为 siRNA 的停泊位点，但并非

自身具有 slicer 酶活性。

随后，他们研究了古菌强烈火球菌（*Pyrococcus furiosus*）的 argonaute-类似蛋白的 X 射线晶体结构（还没有得到全长的真核 Argonaute 结构）。他们发现该蛋白质的 3 个结构域（中间域、PIWI、N 端域）在结构的底部形成“新月”状，PIWI 域位于中间。PAZ 域位于新月的上方，通过一个“蒂”结构域跟新月复合物相连。图 16.34 显示了这一结构，下面的新月复合物形成一个沟，上面有个 PAZ 盖子。这个沟足以容纳双链 RNA，沟内排有一列碱性残基，可与 RNA 底物形成静电桥（electrostatic bridge）。

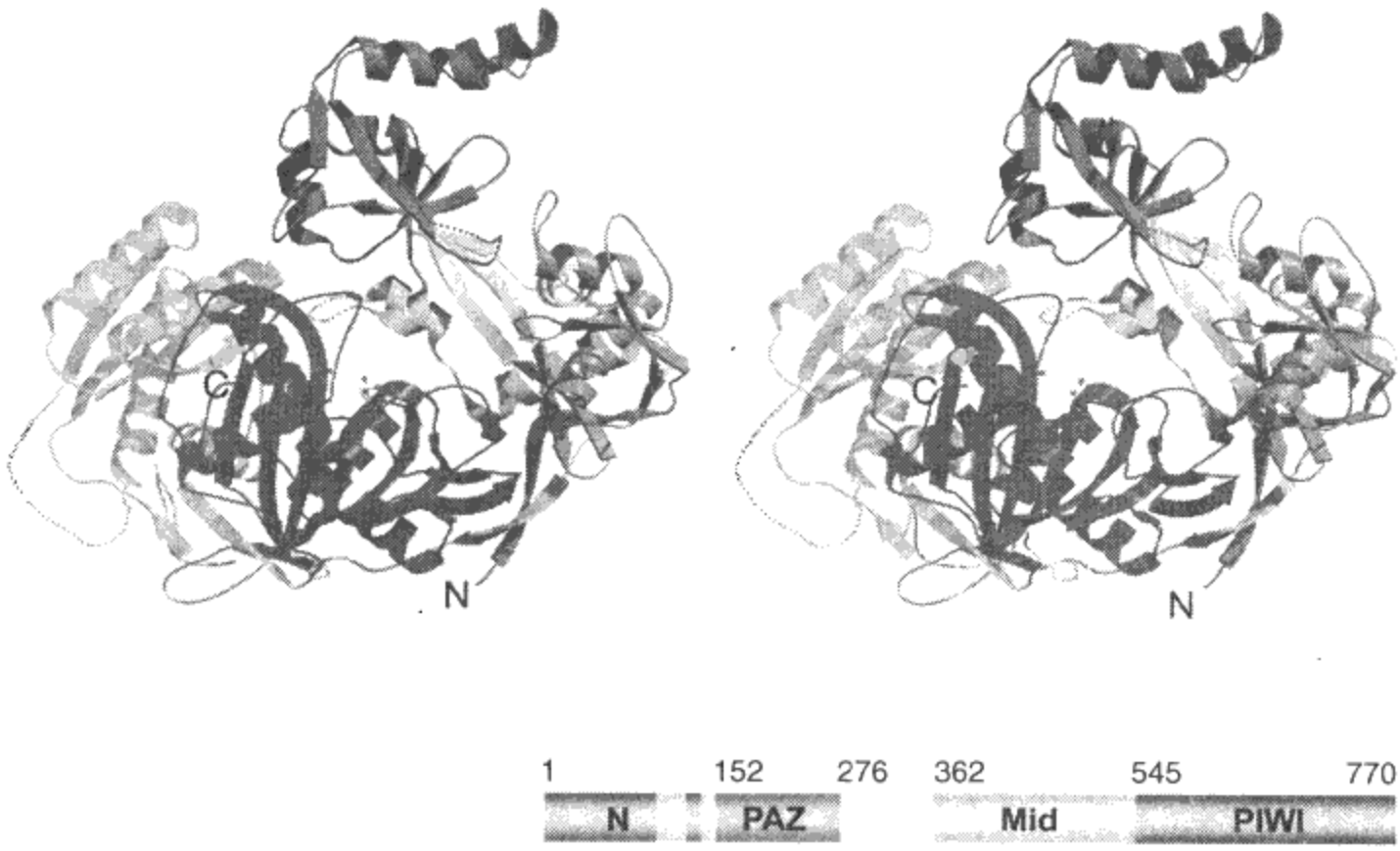


图 16.34　强烈火球菌中 Argonaute 的晶体结构。立体图显示 Argonaute 的结构域，颜色标出与下方示意图一致。注意底部的新月形状，PIWI（紫色）在中间，PAZ 结构域在上面。（*Source*：Reprinted with permission from *Science*，Vol. 305，Ji-Joon Song，Stephanie K. Smith，Gregory J. Hannon，and Leemor Joshua-Tor，“Crystal Structure of Argonaute and Its Implications for RISC Slicer Activity，” Fig. 1，p. 1934，Copyright 2004，AAAS.）

该结构最显著的部分是 PIWI 结构域类似于 RNase H 的一个相似结构域，该域切割 RNA-DNA 杂交链中的 RNA。因此，RNase H 能够识别双链多核苷酸并切割其中的一条链（RNA 链）。除了它们在整个结构上的相似性外，两个蛋白质都有一簇含 3 个酸性氨基的残基（2 个天冬氨酸 1 个谷氨酸）。在 RNase H 中，羧基簇结合 Mg^{2+}，而 Mg^{2+} 在催化 RNA 链切割的反应中起重要作用。这些相似性非常有趣，因为 slicer 具有类似的活性：它必须也识别一个双链多核苷酸（siRNA-mRNA 杂合体）并切割其中的一条链（mRNA）。因此，Argonaute 具有我们所期望的 slicer 的所有特性，一个 PIWI 结构域带有可能负责切割 siRNA-mRNA 杂合体中一条链的位点，另一结构域（PAZ）则结合 siRNA 的一个末端（图 16.35a 和图 16.35b）。

为进一步研究 Argonaute 在哺乳动物中的作用，Hannon 和 Joshua-Tor 及同事对小鼠中的 Argonaute 基因和蛋白质进行了遗传和生化研究。哺乳动物有 4 个 Argonaute 蛋白，分别命名为 Argonaute 1～Argonaute 4。研究人员分别用编码 Argonaute1～3 的基因和萤火虫萤光素酶靶向 siRNA 同时转染细胞。随后利用免疫共沉淀得到 RISC 复合体，体外检测其切割萤光素酶 mRNA 的能力。仅 **Argonaute2**（**Ago2**）具有这一能力。

接下来，研究人员敲除了小鼠的 Ago2 基因，发现所有基因敲除小鼠在胚胎发育阶段就死亡了，并有严重的发育缺陷和发育滞后。造成这一深刻影响的原因是 Ago2 不仅参与 RNAi，同时参与个体的正常发育过程，这一发育过程涉及 microRNA，我们将在本章的后面讨论 microRNA。进一步研究发现，野生型的鼠胚胎成纤维细胞（mouse embryo fibroblast，MEF）表现出正常的 RNAi，而敲除 *Ago2* 基因的小鼠成纤维细胞表现为缺陷的 RNAi，说明 Argonaute2 对于 RNAi 非常重要。

目前所有被引用的研究结果都与 Ago2 具有 slicer 活性这一假说相一致，但仍没有直接证明这一假

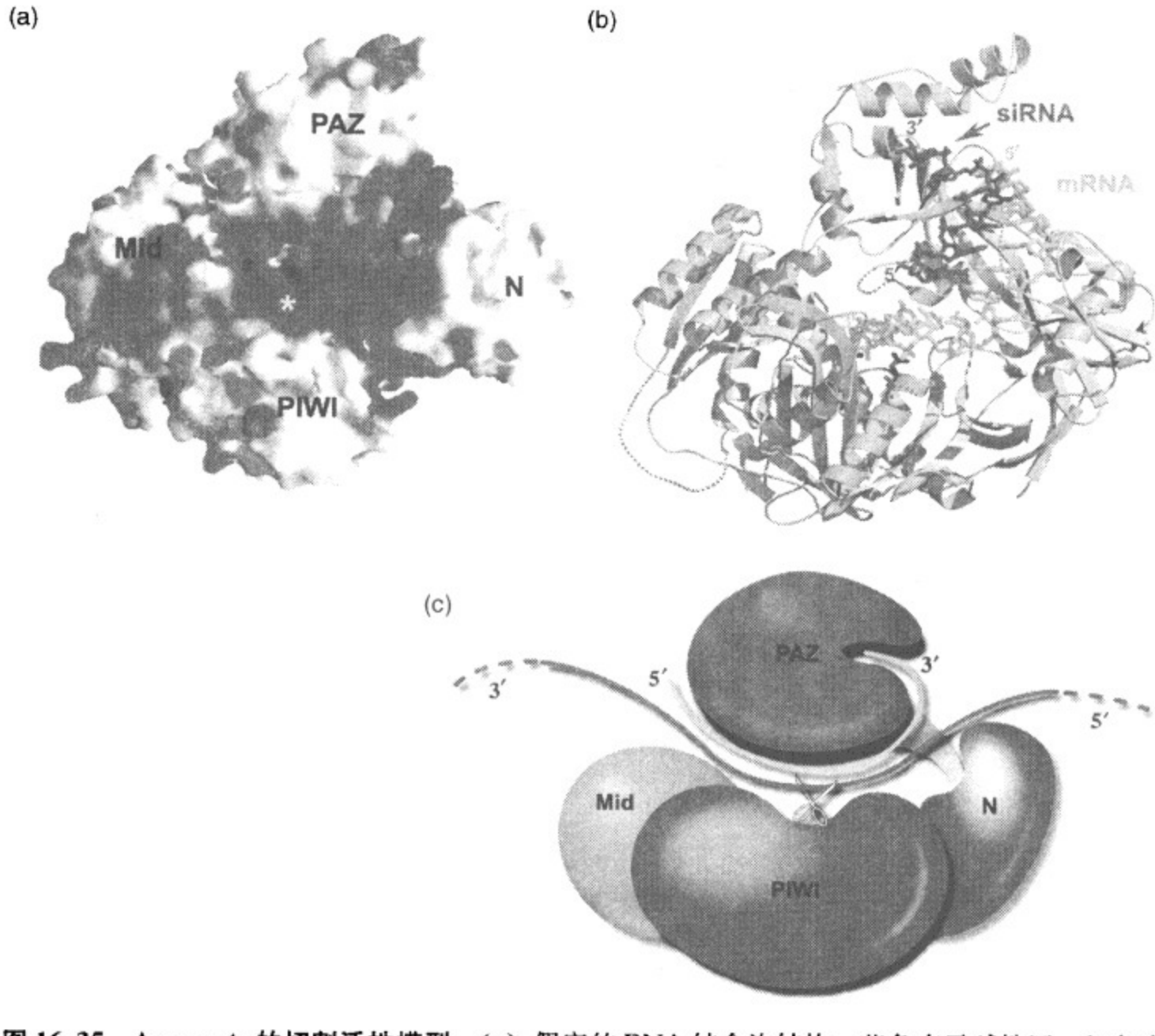


图 16.35　Argonaute 的切割活性模型。（a）假定的 RNA 结合沟结构，蓝色表示碱性区，红色表示酸性区。注意沟内碱性残基占优势。星号表示活性位点。（b）Argonaute 的晶体结构，含有理论上放置的 siRNA 3′端的双螺旋（紫色）和与之碱基配对的目标 mRNA（蓝绿色）。注意，目标 mRNA 的近端和活性位点用 3 个羧基基团（红色）标出，其中 2 个只能部分可见。（c）切割活性模型。siRNA 和目标 mRNA 的杂交体裹在活性位点，至少部分归结于 siRNA 的 3′端与 Argonaute 的 PAZ 结构域相互作用。使目标 mRNA 置于被 slicer 活性位点切割的位置。剪刀表示活性位点。siRNA 为引导 RNA，切割位点位于 siRNA 互补链的中间。（*Source*：Reprinted with permission from *Science*, Vol. 305, Ji-Joon Song, Stephanie K. Smithe, Gregory J. Hannon, and Leemor Joshua-Tor, "Crystal Structure of Argonaute and Its Implications for RISC Slicer Activity," Fig. 4, p. 1436, Copyright 2004, AAAS.）

说。如果 Ago2 确实有 slicer 活性，那么突变其活性作用位点的 3 个酸性氨基酸中的任何一个都应该阻断 RISC 对 mRNA 的切割。Hannon、Joshua-Tor 及同事对其中 2 个关键天冬氨酸残基分别进行突变，通过体内体外实验都证明，任何一个突变体都破坏了 RNAi-mRNA 的切割。综上所述，所有证据都有力地证明 Ago2 是 slicer 酶。

2005 年，Joshua-Tor 及同事很明确地证明了人 Ago2 具有 slicer 活性。他们利用人重组 Ago2 和 siRNA 构建了一个小 RISC 复合体，该复合体可精确切割与 siRNA 互补的底物 RNA。图 16.36 显示，第一个 siRNA（siRNA1）在底物 RNA（S500）3′端约 140nt 处切割，产生约 140nt 的 3′产物和约 360nt 的 5′产物。第二个 siRNA（siRNA2）在 S500 RNA 5′端约 180nt 处切割，产生约 180nt 的 5′产物和约 320nt 的 3′产物。在缺少 siRNA 或 Mg^{2+} 时都无切割产物，表明二价金属离子是 slicer 所必需的。

要产生 mRNA 切割，必须形成一个具有催化活性的 RISC 复合体（图 16.37）。我们已经看到 Argonaute 蛋白含有 RISC 的 slicer 活性位点，同时必须有一个单链 siRNA 作为向导选择性降解 mRNA。所以，Ago2 加上 siRNA 可组成一个小 RISC，至少在

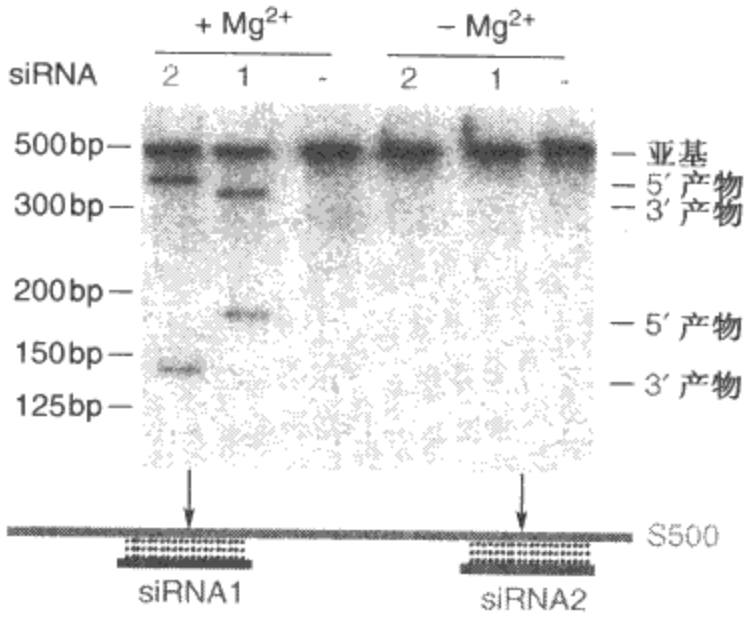


图 16.36　Ago2 和 siRNA 组成具有体外切割活性的小 RISC 复合体。Joshua-Tor 及同事将重组人 Ago2 蛋白（在大肠杆菌中生产）分别与 2 个 siRNA 混合，这 2 个 siRNA 可特异地作用于一个 500nt 目标 RNA 的 2 个不同位点，如图下部所示。在 Mg^{2+} 有或无条件下，分别加入标记的目标 RNA，如图顶部所示，所用 siRNA 也在顶部标出。最后，电泳检测标记的 RNA 产物。可见，切割活性依赖于 Mg^{2+} 和 siRNA。这 2 个 siRNA 作用于 mRNA 产生了不同的产物，根据 mRNA 与 siRNA 互补位置所预测的产物大小与之相同。(*Source*: Reprinted from *Nature Structural & Molecular Biology*, vol 12, Fabiola V Rivas, Niraj H Tolia, Ji-Joon Song, Juan P Aragon, Jidong Liu, Gregory J. Hannon, Leemor Joshua-Tor, "Purified Argonaute2 and an siRNA form recombinant human RISC," fig. 1d, p. 341, Copyright 2005, reprinted by permission from Macmillan Publishers Ltd.)

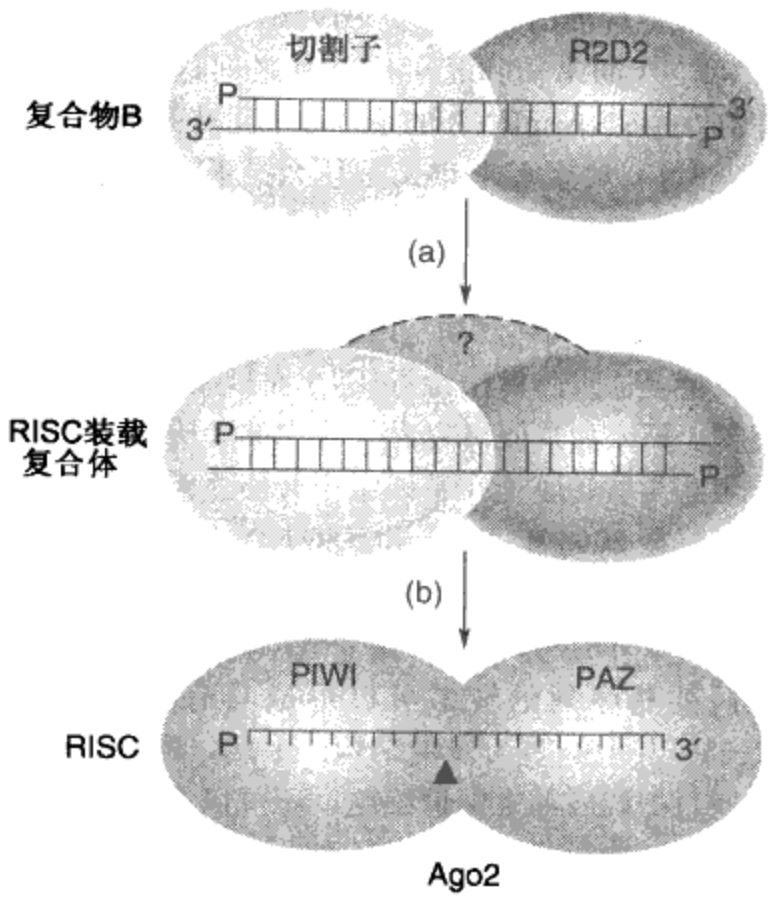


图 16.37　RISC 复合体的装配。（a）复合体 B 由 siRNA、Dicer（黄色）和 R2D2（蓝色）组成，在加入其他蛋白条件下转变为 RLC，加入的蛋白质可能是 Armitage（粉红色）。(b) RLC 将 siRNA 送给 Argo2（橙色）。在这一过程中，需要对 siRNA 双链进行解链，除去伴随链，留下引导链与 Argo2 结合，该引导链的 5′端位于 PIWI 结构域中，而其 3′端位于 PAZ 结构域中。活性位点（紫色）位于 siRNA 的第 10～11 碱基。在体内可能还有其他的蛋白质参与 RISC，但本图显示的仅是体外观察到的小 RISC。

哺乳动物中是这样。但这种复合体不会直接形成，需要有 **RISC 装载复合体**（RISC loading complex，**RLC**）将 siRNA 加载到 Ago2 中，RLC 显然能分离 siRNA 的双链，并选择其中的引导链加载到 RISC 中。

在果蝇中，由 RLC 转变成 RISC 需要蛋白 Armitage 参与，Armitage 可能作为解螺旋酶使 siRNA 解链，尽管在 RLC 中已经含有单链 siRNA，Armitage 仍然是必需的（如果 RLC 不能分离两条 siRNA 链，那么在 RISC 形成后，终止链分离会立即发生，因为很少发现 Ago2 与双链 siRNA 结合在一起）。RLC 的组成尚不清楚，推测至少含有 Dicer、Dicer 相关蛋白 R2D2、siRNA 以及 Armitage。

R2D2 的作用是什么呢？双链 siRNA 的形成不需要它，因为体外实验证明，在无 R2D2 条件下 Dicer 可以高效完成双链 siRNA 的形成。然而，凝胶阻滞实验和蛋白质-RNA 交联实验均表明，一旦 Dicer 产生了 siRNA，Dicer 单独不能保持与 siRNA 的结合，但 Dicer 加上 R2D2 就可以保持与 siRNA 的结合。此外，R2D2 含有 2 个 dsRNA 结合域，该结合域的突变使 Dicer-R2D2 复合体不能结合双链 siRNA。因此，R2D2 是 RLC 所必需的，它可以在 Dicer 产生了 siRNA 和 siRNA 被运送至 RISC 这段时间内看住 siRNA。

什么因素决定了 siRNA 的哪条链为**引导链**（guide strand），哪条链为被抛弃的**伴随链**（passenger strand）呢？这一区分发生在复合体 B 中，该复合体形成于 RLC 之前，含有分别结合于双链 siRNA 一个末端的 Dicer 和 R2D2。这两个蛋白质的结合不对称，Dicer 结合 siRNA 不稳定的末端（该末端的碱基对容易分开）。5′端结合了 Dicer 的链将成为引导链。

对 siRNA 和 Argonaute 样蛋白之间的复合体 X 射线晶体结构研究表明，siRNA 引导链的 5′磷酸末端结合于 Argonaute 的 PIWI 口袋，其 3′端结合于

PAZ域，这种结合正好将 Argonaute 的活性位点置于 siRNA 的第 10 和 11 位碱基之间，因此 mRNA 正好在 siRNA-mRNA 的杂合链中间被剪切。

RNAi 的生理学意义是什么？正常情况下，真核生物细胞不存在双链 RNA。但是当受到某些外源 RNA 病毒感染时就会发生，这些 RNA 病毒要通过双链 RNA 中间体复制。因此，RNAi 的一个重要功能可能是通过降解病毒 mRNA 而抑制 RNA 病毒的复制。Fire 和其他研究者发现 RNAi 所需的基因也是阻止某些转座子在基因组中转座所需要的。事实上，早在 2003 年 Titia Sijen 和 Ronald Plasterk 就发现线虫精细胞中 Tc1 转座子的转座活性受 RNAi 沉默。哪些 dsRNA 触发了这一 RNAi 呢？转座子末端反向重复序列的转录可以形成具有茎环结构的 RNA，其中的茎为双链结构。因此，RNAi 不仅能够通过抵抗病毒来保护细胞，还可以通过抑制转座子来保证精细胞基因组的完整性。

RNAi 也可以沉默转基因及其基因组的同源基因。双链 RNA 是如何从转基因中产生的呢？与正常基因不同，好像转基因的两条链都发生了转录。这种对称转录可产生足够的双链 RNA 以激活 RNAi。

除了其天然功能外，RNAi 对分子生物学家来说是一个极大的恩赐，因为 RNAi 让他们可以随心所欲地沉默基因，只要简单地引入目标基因相对应的双链 RNA 即可。这一操作比实验室产生基因敲除生物的过程（如第 5 章所述）要方便得多。当然这一技术也受到生物技术产业的关注，RNAi 展现了潜在的利润。我们知道许多基因过于活化时会产生破坏作用。例如，许多**原癌基因**（oncogene），在肿瘤细胞中具有超级活性，使肿瘤细胞生长失控。直接抵抗原癌基因的 RNAi 可以控制原癌基因的活性，进而恢复对癌细胞的生长控制。

尽管对 RNAi 非常乐观，但谨慎些更保险，因为 2004 年以来的数据显示，RNAi 不如想象的那样精准特异，与触发 dsRNA 不完全配对的基因也受到一定程度的抑制。我们还不知道这一非特异性是否会严重影响 RNAi 在科学研究和医学应用中的效率。

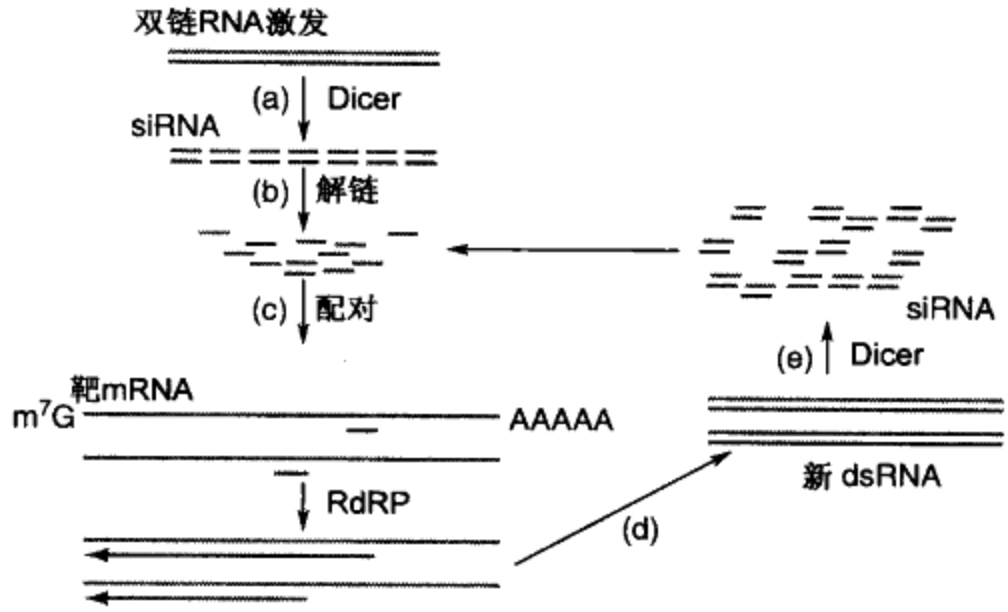
此外，如果科学家们要运用 RNAi 技术研究人类基因的功能或是治疗疾病，他们还得考虑另外一个事实：与蛔虫和果蝇不同，靠引入 dsRNA 到哺乳动物细胞所引发的 RNAi 是瞬时的。有一个解决办法是给哺乳动物细胞转染某种基因，该基因编码具有反向重复序列可形成发夹的 RNA。这些基因以发夹结构的形式持续提供 dsRNA，使 RNAi 持续进行。截至 2004 年，研究人员已经建立了一个编码**短发夹 RNA**（short hairpin RNA，**shRNA**）的基因文库，可以靶向作用于近 10 000 个人类基因。这对于科学研究甚至人类疾病的治疗都将是十分有价值的资源。

小结 细胞遭遇 dsRNA 侵入时会发生 RNAi，这些 dsRNA 可能来自病毒、转座子或转基因（或者实验加入的 dsRNA）。这种触发 dsRNA 由一个类似 RNase III 的 Dicer 酶降解成 21～23nt 的片段（siRNA）。双链 siRNA 和 Dicer 以及 Dicer 相关蛋白 R2D2 组成复合体 B。复合体 B 将 siRNA 传递给 RISC 装载复合体 RLC，RLC 可能将双链 siRNA 分成两个单链 siRNA，同时将引导链转运给 RNA 诱导沉默复合体 RISC，RISC 含有称为 Argonaute2（Ago2）的蛋白质。在 Ago2 PIWI 结构域的活性位点，siRNA 的引导链同目标 mRNA 碱基互补配对，其中 PIWI 结构域类似于 RNase H 酶，也被称为 slicer。slicer 在目标基因同 siRNA 互补配对区域的中间切割 mRNA。在一个依赖 ATP 的步骤中，被切割的 mRNA 从 RISC 上释放出来，以便接受一个新的待降解 mRNA。

siRNA 的扩增 很难解释 siRNA 在一些生物体如植物和线虫中的高灵敏性，仅少数 dsRNA 分子就可以调动完全沉默一个基因的过程，而且不仅只在一个细胞中，而是在整个生物体，甚至其后代中发生。这一现象表明基因沉默过程是催化性的。确实，Dicer 从 dsRNA 和目标 mRNA 产生出许多 siRNA 分子，但这仍不足以解释 RNAi 的巨大威力。2001 年，Bruce Paterson 及同事揭开了这一谜底，他们发现细胞利用 **RNA 指导性 RNA 聚合酶**（RNA-directed RNA polymerase，**RdRP**）以反义 siRNA 为引物扩增 siRNA 的拷贝，如图 16.38 所示。

首先，Paterson 及同事证明 20pmol/L（20×10^{-12}）的 GFP mRNA 所对应的触发 dsRNA 就可在果蝇胚提取物中产生 siRNA，同时，GFP mRNA 被降解但非目标的萤光素酶 mRNA 不受影响。可见 RNAi 可在体外复制。接下来，他们发现，利用 GFP dsRNA 所合成的 siRNA 可以体外诱导 RNAi 的产生。但是，合成的 siRNA 必须具有 3′-OH 基团才能诱导 RNAi，表明 siRNA 作为引物引导 RdRP 扩增 mRNA 产生 dsRNA。

为验证这一假说，他们用 UTP［^{32}P］标记 siRNA，检测被标记的 siRNA 是否以引物形式进入全长 dsRNA。同时，他们检测 dsRNA 是否同目标 mRNA 相应的长度一致（GFP 和萤光素蛋白酶分别为 716bp 和 1652bp）。结果非常明确，如图 16.39a 所示，以 mRNA、反义 RNA 和 dsRNA 为模板时，标记 siRNA 都可掺入全长 RNA，但无模板或模板非同源时都不能掺入。此外，标记的产物可以抵抗 RNase T1 和 RNase I 的降解，表明它们是双链的（图 16.39b）。



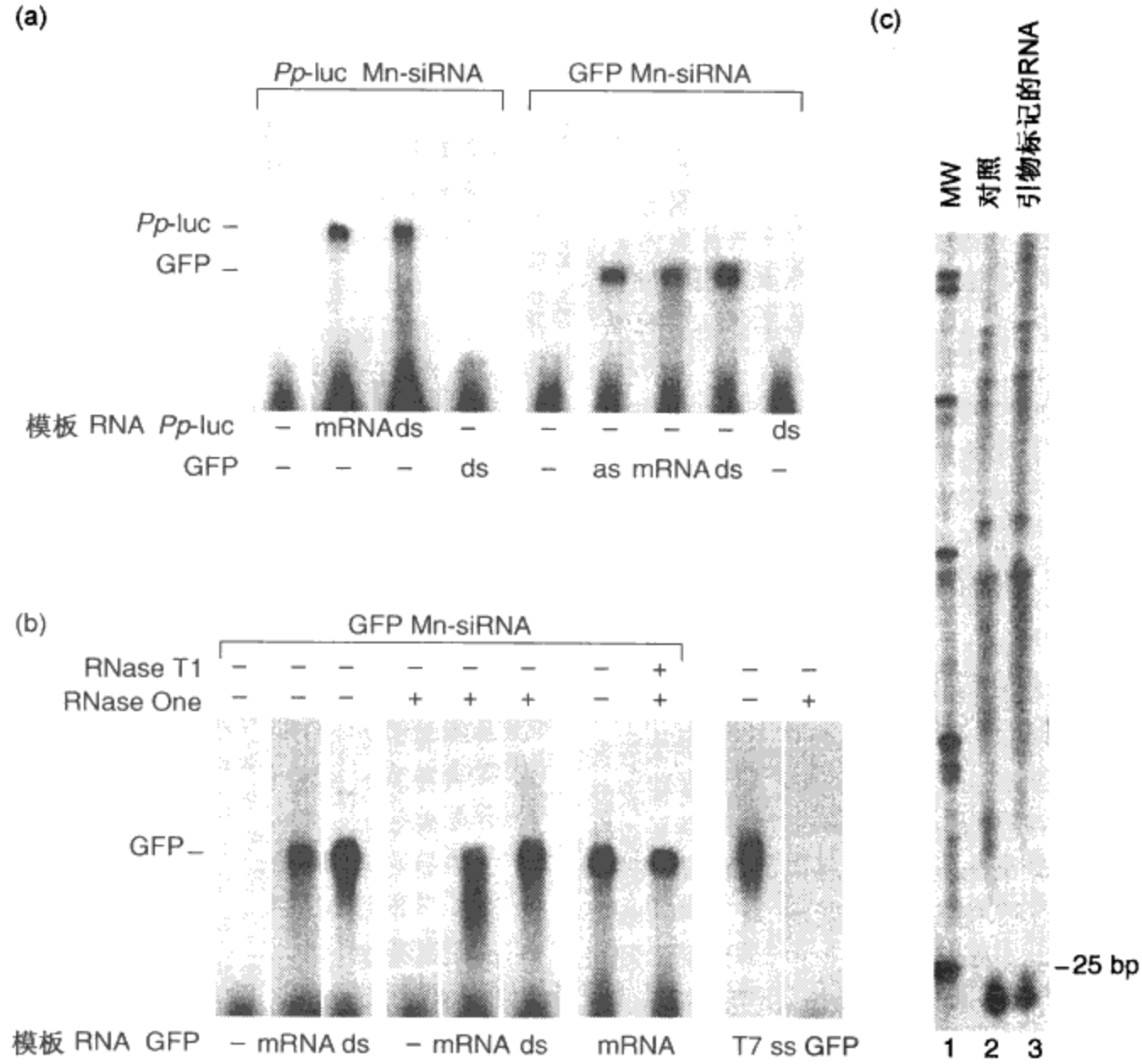


图 16.38　siRNA 的扩增。(a) Dicer 剪切触发 dsRNA 成为 siRNA；(b) siRNA 的无义链与靶 mRNA 杂交；(c) RdRP 利用 siRNA 的无义链为引物，靶mRNA为模板合成长的无义链；(d)(c) 中产生的是新的触发 dsRNA；(e) Dicer将新的触发 dsRNA 剪切成更多新的 siRNA，这些 siRNA 又引发新一轮的扩增。(*Source*: Adapted from Nishikura. *Cell* 107 (2001) f. 1, p. 416.)

图 16.39　siRNA 引发目的 dsRNA 的合成。(a) 全长触发 dsRNA 的生成。Paterson 及同事把人工合成的标记 siRNA 分别在无 RNA (–)、目标 mRNA、触发 dsRNA、非同源 RNA 条件下（图底部所示），加入果蝇胚胎提取物中，凝胶电泳检测产物。分别使用萤光素酶 (Luc) 或绿色荧光蛋白 (GFP) 的 siRNA（图顶部所示）。Luc 或 GFP 的全长 mRNA 的位置在图左边标出 (ds＝双链 RNA，as＝反义 RNA)。同源 RNA、反义 RNA 或 dsRNA 分别作为合成全长 RNA 的模板。(b) 全长 dsRNA 产物对 RNase 的抗性。(a) 中标记的 GFP 的 RNA 用 RNase T1 或 RNase I 处理后，凝胶电泳检测。模板在图底部标出。(c) 全长 dsRNA 产物加工成 siRNA。纯化 (a) 中 GFP 的全长双链 RNA，与果蝇胚胎提取物孵育，凝胶电泳检测（泳道 3）。泳道 1 为分子质量标准，泳道 2 为对照，已知要加工成 siRNA 的 GFP 全长双链 RNA。siRNA 条带在 25bp 之下。(*Source*: Reprinted from *Cell* v. 107 Lipardi et al. Fig. 4, p. 301, © 2001 with permission from Elsevier Science.)

Paterson 等纯化了细胞提取物中由标记 siRNA 产生的标记 dsRNA，并将其回加到细胞提取物中。图 16.39c 显示，提取物（可能是 Dicer）加工这种 RNA 而产生新的 siRNA。延长孵育时间后又显示了降解的证据：首先全长 dsRNA 中出现了标记信号，在随后的 10min 内 dsRNA 开始消失，siRNA 的量再次增加。因此应存在一个机制扩增被加入的 dsRNA，这可以解释 RNAi 的强大效力。该反应的第一轮依赖于以反义 siRNA 为引物以 mRNA 为模板的引发。该模式可以解释 Fire 及同事早先的发现：dsRNA 的无义链而不是有意义链的改变阻断了 RNAi。这一模式同早期在番茄中发现的 RdRP 和在真菌、线虫以及其他植物中出现的同源基因是一致的，它们是 RNAi 有效进行所必需的。

令人鼓舞的是，Fire 及同事在线虫进行的类似实验中得到了相似的结果。他们用超量表达触发 dsRNA 的细菌喂养线虫，以标记的有义链为探针进行 RNase 保护实验，检测线虫中的反义 siRNA 链。实验中采用了两种不同的触发 dsRNA，发现两种情况下都有大量新 siRNA 产生。此外，他们在触发 RNA 区域外发现了一些次级 siRNA（secondary siRNA）。显然，这些次级 siRNA 只与触发序列上游的 mRNA 区域相对应。这一结果同 RdRP 的作用特点相一致，因为 siRNA 引导 RNA 合成的反应向着 mRNA 的 5′端（即上游）进行。因此，次级 siRNA 的发现也佐证了 RdRP 以目标 mRNA 为模板扩增 siRNA 的假说。

小结 siRNA 的扩增是在 RNAi 过程中，反义 siRNA 杂交到目标 mRNA 上，在 RdRP 的作用下扩增得到全长反义 RNA，这些 dsRNA 被 Dicer 消化剪切成新的 siRNA 片段。

RNAi 机制在异染色质形成和基因沉默中的作用

在 2002 年，有越来越多的证据显示，在异染色质形成和基因沉默以及 RNAi 自身中都涉及 RNAi 体系。Shiv Grewal、Robert Martienssen 及同事在裂殖酵母（*Schizosaccharomyces pombe*）中分别删除了编码 Dicer、Argonaute 和 RdRP（分别为 *dcr1*、*agol* 和 *rdp1*）的 RNAi 基因，发现所有突变体在沉默插入着丝粒附近的转基因方面是缺陷的，而正常情况下着丝粒附近插入的转基因的表达是受抑制的。也就是说，转基因在 RNAi 突变体中有活性。注意，此处没有加入针对转基因的 dsRNA，因此 RNAi 不是直接参与对转基因的沉默。

同时研究人员利用 Northern 印迹检测了着丝粒的重复 DNA 序列（*cen3* 序列）在野生型和突变体中是否得到转录。结果显示，在野生型细胞中无转录物，而在 3 个 RNAi 突变体中发现有大量 *cen3* 转录物。利用 RNA 斑点杂交进行更详细的研究显示，在野生型和突变体细胞中有 *cen3* 的反转录物，而正向转录物仅在突变体细胞中出现。此外核连缀转录显示了相同的转录模式，即正向转录仅发生在突变体中。可见，对 *cen3* 转录物浓度的控制发生在转录水平上而不是转录后水平上。

接着，研究人员利用染色体免疫沉淀技术（ChIP）检测特异核心组蛋白的甲基化现象，所用抗体为抗组蛋白 H3 的抗体，该 H3 的第 4 和第 9 位赖氨酸均发生了甲基化。其中第 4 位赖氨酸甲基化同基因的活化有关，而第 9 位赖氨酸甲基化与异染色质和基因失活相关。与我们前面讨论的一样，野生型细胞着丝粒区域的第 4 和第 9 位赖氨酸都发生了甲基化，而 3 个 RNAi 突变体的着丝粒 H3 组蛋白表现出异常甲基化：第 4 位赖氨酸的甲基化水平较高，而第 9 位赖氨酸的甲基化水平很低。这一甲基化组合与在 *ura4*$^+$ 转基因菌株中发现的形式一样，该转基因位于最外侧着丝粒区域（outermost centromere region，*otr*）；在野生型细胞中第 9 位赖氨酸甲基化水平较高，而在 3 个 RNAi 突变体中该赖氨酸甲基化水平受到极大地抑制。

RNAi 与组蛋白的甲基化，以及进而引起着丝粒处异染色质化有关吗？如果有关的话，可以预期至少有一些 RNAi 蛋白与着丝粒染色质相互作用，同时还可预期应该存在与着丝粒 RNA 相应的 siRNA。Martienssen 及同事确实发现了 RNAi 机器中的 Rdp1 部件与着丝粒染色质相结合。同时，B. J. Reinhard 和 David Bartel 也提供了支持该假说第二个推论的有力证据，他们从野生型细胞中克隆到了 Dicer 的产物，显示所有 12 个克隆都源于着丝粒区域的转录物。

因此，在着丝粒处至少发现了 RNAi 机器的一个部件，并且 siRNA 来源于着丝粒转录物。根据这些结果和其他资料，Martienssen 及同事提出，着丝粒处异染色质化的沉默与 RNAi 有关（图 16.40）。具体讲，丰富的 *otr* 区的反向转录物与 RNA 聚合酶或 RdRP 偶尔转录的正向转录物碱基互补配对，形成触发 dsRNA。然后 Dicer 消化 dsRNA 产生 siRNA，而 siRNA 同 **RITS 复合体**中的 Argonautel 蛋白（**Agol**）结合（RITS 复合体是指 RNA 诱导的转录基因沉默起始物，RNA-induced initiator of transcrional gene silencing）。RITS 复合体然后吸引 **RDRC 复合体**（RNA 指导的 RNA 聚合酶复合体，RNA-directed RNA polymerase complex）中的 RdRP，RDRC 可以扩增双链 siRNA。通过与该 DNA 直接配对或与其转录物配对，siRNA 将 RITS 护送到基因组的相应位点上。然后，RITS 召集组蛋

白 H3 Lys9 甲基转移酶。一旦 Lys9 甲基化，它又召集 Swi6，Swi6 是异染色质化所必需的。可能还需要其他蛋白质，但最终结果是异染色质化扩散到着丝粒的 *otr* 区。无论是什么机制，似乎都是高度保守的，因为哺乳动物外周着丝粒的异染色质结构也涉及组蛋白 H3 Lys9 的修饰以及一些 RNase 敏感物质，这些 RNase 敏感物极有可能是一种或多种 RNAi 的中间产物。

Grewal 及同事注意到，在一些位点也发现了类着丝粒（centromere-like）序列，如交配型沉默区，该区离着丝粒较远，但仍受异染色质化而沉默。在另外的实验中，研究人员发现交配型沉默区的异染色质化的起始需要 RNAi 机器，而维持则不需要。Swi6 显然足够异染色质的维持。

RNAi 机器在着丝粒事件中的作用不仅仅局限于低等生物。2004 年，Tatsuo Fukagawa 及同事报道了在鸡-人杂交细胞系中的检测结果，该细胞系仅含人类的第 21 号染色体。他们让杂交细胞系中 Dicer 基因的表达受到四环素的抑制，然后观察细胞的变化，尤其是观察在四环素阻遏 Dicer 表达的情况下人类第 21 号染色体的变化。Dicer 缺失最明显的影响是在大约 5d 后这些细胞发生死亡。

此外，这些细胞表现出与着丝粒相关的特异的病理学特征，即在有丝分裂期间的姊妹染色单体发生了异常分离。像 RNAi 缺陷的酵母细胞一样，在这些脊椎动物的细胞中，人类 21 号染色体的着丝粒重复序列的转录物不正常地堆积。他们也发现了一些着丝粒蛋白的不正常定位，但不是全部着丝粒蛋白。在着丝粒发生的这些现象可能是由于 Dicer 的缺失所致，继而导致细胞分裂失败，以及细胞死亡。

我们认为图 16.40 所示的发生在裂殖酵母着丝粒区域的事件有助于解释高等生物细胞中产生的上述结果。但哺乳动物似乎缺少 RdRP。因此出现在哺乳动物着丝粒处的 dsRNA 肯定是由该区域 DNA 的双向转录所产生，或基因组某处同源区域的双向转录所产生。

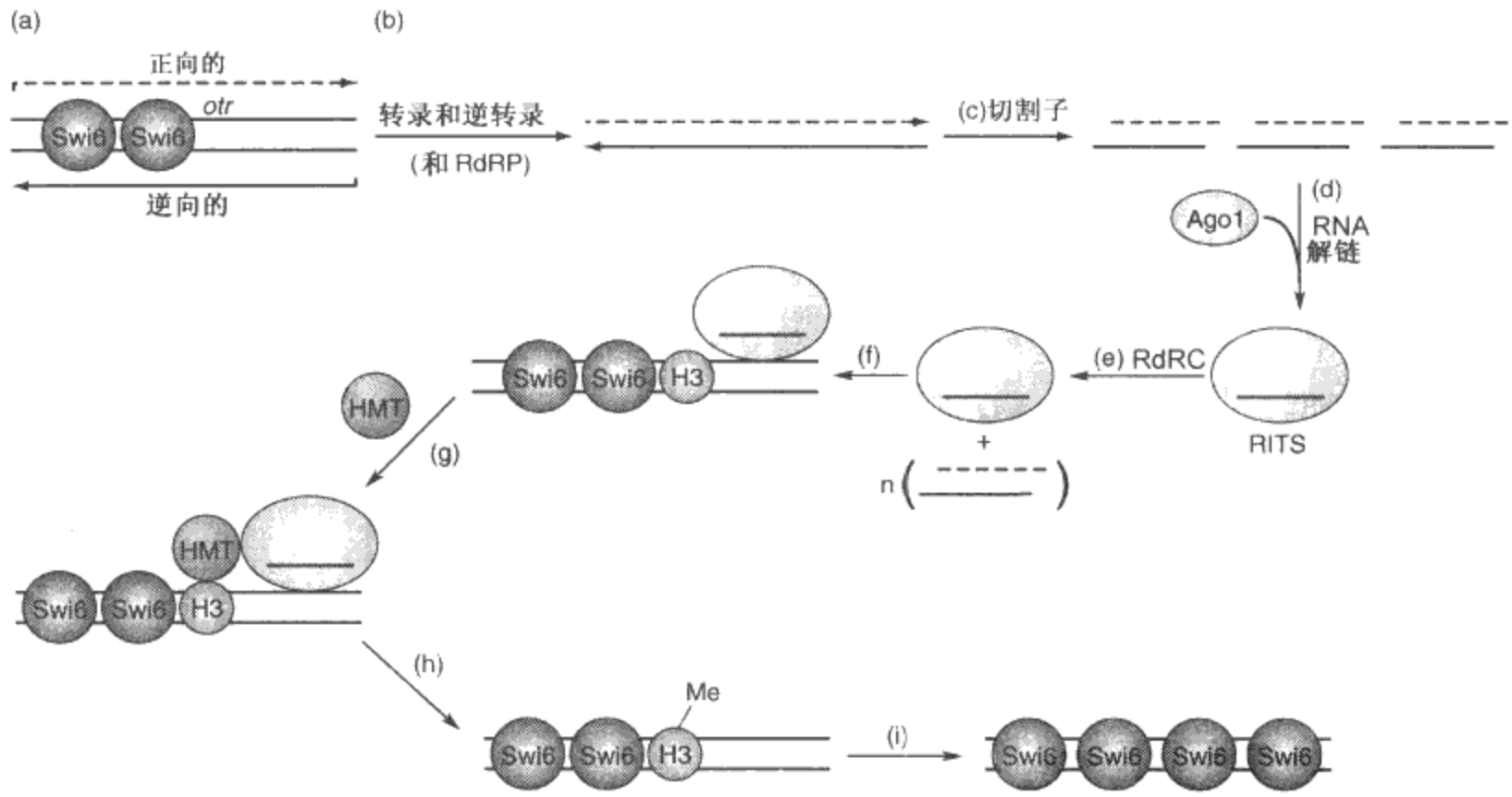


图 16.40　在 *S. pombe* 的着丝粒异染色质化中 RNAi 机器的参与模型。（a）着丝粒最远端区（*otc*）恒定地产生反向转录物，正向转录物可能也产生，但水平很低（检测不到）；（b）转录和反向转录之后（或反向转录和 RdRP 作用之后），得到双链 DNA（dsRNA）；（c）Dicer 将 dsRNA 切割成 siRNA；（d）Ago1（黄色，可能与其他蛋白质在一起）与单链 siRNA 结合产生 RITS；（e）RDRC 中的 RdRP 扩增 siRNA，产生双链 siRNA；（f）RITS 通过其 siRNA 与 *otr* 结合，或是通过直接与 DNA 作用或是与该 DNA 的转录物作用；（g）RITS 将组蛋白甲基转移酶（HMT，绿色）吸引到 *otr* 上；（h）HMT 将组蛋白 H3 的 Lys9 甲基化，当然，该组蛋白是核体的一部分，为简单起见，此处未显示核体；（i）这种甲基化吸引了更多的 Swi6（红色），使异染色质化扩展。

裂殖酵母同植物和哺乳动物异染色质化的另一主要区别是，除组蛋白甲基化外，动植物还发生了 DNA 的甲基化。甲基基团加到双链 CpG 序列的 C 上，它们有助于吸引那些能够诱导染色体发生异染色质化的蛋白质。还有，dsRNA 的出现对 RNAi 机器的募集具有重要的作用，RNAi 机器可刺激 DNA

甲基化。

这一机制的一个显著优点是它的永久性。一旦DNA CpG序列双链中的C被甲基化，这种甲基化便会从一代细胞遗传给下一代。DNA复制完成后，一条链上甲基化的C可确保相反链中新掺入的C也发生甲基化。虽然甲基化是永久的，但不是真正的遗传学改变，可能是一种碱基变为另--个碱基（如C变为T）。我们称之为DNA的**表观修饰**（epigenic modification）。它同遗传学的改变一样重要，因为它可以导致基因的沉默甚至可以导致染色体整个区段的异染色质化。

RNAi可能在哺乳动物X染色体的失活中起作用。雌性哺乳动物的每个细胞中，一条X染色体被异染色质化而失活，避免了过多X染色体产物对机体的毒害作用。在X染色体失活过程中，第一步是组蛋白H3 Lys9的甲基化，该甲基化在*Xist*基因座的非编码转录物出现后立即产生。我们知道，*Xist*的转录受反义RNA、*Tsix*和*Xist*启动子甲基化控制。*Xist*和*Tsix*的转录物在同一细胞同时出现时会触发RNAi系统，RNAi系统召集组蛋白甲基化酶，进而引起异染色质化的形成。

2004年Kevin Morris及同事发现，哺乳动物基因也能被RNAi机器所沉默，正像我们在植物和哺乳动物中所看到的异染色质化一样。这种基因沉默涉及DNA甲基化。此外，与正常RNAi不同的是，这种沉默所涉及的siRNA直接作用于基因的调控区，而不是编码区。

Morris及同事构建了一个GFP报告基因，由人类延伸因子1α基因（EF1A）的启动子-增强子区控制。用含有GFP报告基因的猫科动物免疫缺陷病毒（feline immunodeficiency virus，FIV）转导人细胞，使该报告基因及其调控区整合到人基因组中。FIV载体也使核膜对siRNA具有通透性，否则siRNA不能被哺乳动物细胞核吸收。

因为实验所用siRNA直接对应基因的调控区而不是编码区，可以预测它不会导致mRNA降解或阻断翻译，但可阻断转录。事实正如Morris及同事所发现的那样。他们利用实时定量RT-PCR（第4章）证明，EF52 siRNA转导细胞后几乎所有GFP转录物都消失了，而EF52 siRNA靶向GFP融合基因的调控区。相反，靶向GFP mRNA编码区的siRNA均导致GFP转录物浓度下降78%（图16.41a）。

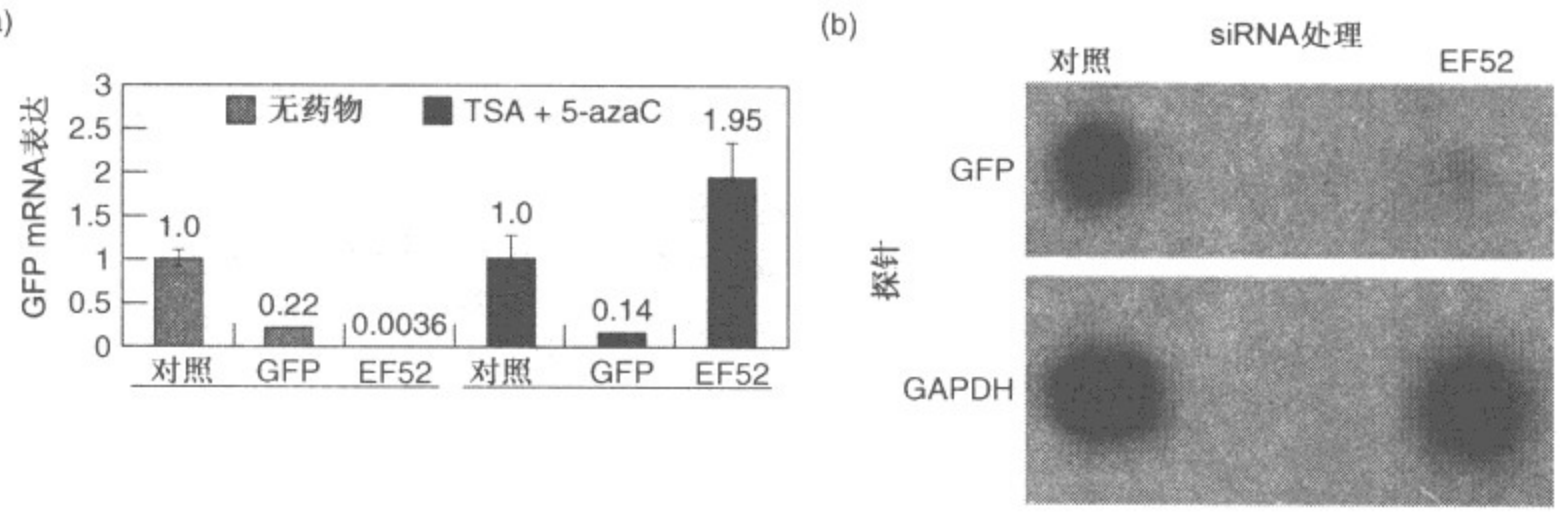


图16.41　siRNA靶向EF1A基因的调控区域而造成的基因沉默。（a）实时定量PCR实验检测人细胞中GFP mRNA，GFP基因由EF1A基因的启动子-增强子区驱动。用含GFP基因重组子的FIV转导细胞，然后在有或无TSA和5-azaC药物的情况下加入siRNA。然后用实时定量PCR测定GFP mRNA的浓度。柱状图（及其相应的量）表示加入不同siRNA的结果：对照没有siRNA、靶向GFP基因编码的siRNA（GFP）、靶向EF1A基因调控区的siRNA（EF52）。（b）核连缀转录实验。从转有EF1A-GFP质粒的细胞中分离细胞核，加入EF52 siRNA或不加入siRNA（对照）。标记的连缀转录物分别同GFP DNA和GAPDH DNA进行斑点杂交，如左侧所示。在转录水平上，EF52 siRNA沉默GFP基因，但不影响GAPDH基因。（*Source*：Reprinted with permission from *Science*，Vol. 305，Kevin V. Morris，Simon W.-L. Chan，Stever E. Jacobsen，and David J. Looney，"Small interfering RNA-Induced Transcriptional Gene Silencing in Human Cells，" Fig. 1，p. 1290，Copyright 2004，AAAS.）

哺乳动物发生转录沉默的共同特征是组蛋白和DNA的甲基化，故Morris及同事分别检测了曲古抑菌素（trichostatin，TSA）和5-氮胞苷（5-azacytidine，5-azaC）对转录沉默的影响。TSA和5-azaC可分别抑制组蛋白和DNA的甲基化。这些药物完全逆转了由EF52 siRNA所造成的基因沉默，但对GFP编码区siRNA所引起的基因沉默没有影响。这些结果支持了DNA和（或）组蛋白甲基化参与由EF52引起的基因沉默的假说。

为了检测由EF52引起的基因沉默是否发生在

转录水平上，Morris 及同事进行了核连缀转录实验（第 5 章）。图 16.41b 显示，EF52 确实大大降低了 GFP 的转录量，而无关的 GAPDH（磷酸甘油醛脱氢酶，glyceraldehyde-phosphate dehydrogenase）转录物不受影响。

为了验证在转录沉默中基因调控区的 DNA 是否发生了甲基化，他们又利用识别位点包含 CpG 的限制性内切核酸酶 *Hin*P1I 进行实验。若序列中的 C 未甲基化，*Hin*P1I 可以酶切，否则不能酶切。在 EF1A 的调控区有一个 *Hin*P1I 酶切位点，如果该位点发生了甲基化，就可免受 *Hin*P1I 酶切，利用该位点两侧的引物进行 PCR 会得到扩增产物。反之，若该位点未甲基化，*Hin*P1I 会对其酶切而得不到 PCR 产物。

图 16.42 显示了这一实验的结果。泳道 1 的对照显示，含有 *Hin*P1I 位点的质粒在体外甲基化后的确实产生了 PCR 产物，甚至在经过 *Hin*P1I 处理后仍然可产生 PCR 产物。泳道 2 和 3 分别表示来自无关 siRNA 和 GFP 编码区相应 siRNA 转导后细胞的 DNA 对照。泳道 4 表示转导 EF52 siRNA 后的结果。顶部一行表示 DNA 一定发生了甲基化，因为它被保护而免受 *Hin*P1I 的酶切，结果出现一条 PCR 产物。然而底部一行表示阻断甲基化的药物 TSA 和 5-azaC 使 *Hin*P1I 位点被酶切，因而无 PCR 产物。

至今所提到的利用 FIV 转导细胞的实验中，FIV 将 EF1A 基因整合到人基因组中，而不是在其天然存在的位置上。为检测 siRNA 对内源人类基因的沉默机制，Morris 及同事进行了与图 16.41 和图 16.42 类似的实验，但所用细胞在 MPG 的作用下对 siRNA 具有通透性，MPG 指的是含有 HIV-1 转膜肽以及一个与之连接的 SV40 病毒核定位信号的融合多肽。在这些实验中，没有 EF1A 进入细胞，仅有内源基因存在，并且内源基因被 EF52 siRNA 沉默（虽然不像前述实验发生那么强的基因沉默）。如同上述情况一样，基因沉默伴随有 DNA 的甲基化，但这一沉默可被甲基化抑制物所阻断。

小结 RNAi 机器参与酵母着丝粒和沉默交配型区域的异染色质化，也参与其他生物的异染色质化。在裂殖酵母着丝粒的最外侧区域无义链发生了转录。偶尔发生的正向转录或由 RdRP 所形成的正向转录物同反向转录物碱基配对，引发 RNAi，继而召集组蛋白甲基转移酶，该酶使组蛋白 H3 Lys9 甲基化，而甲基化的组蛋白 H3 又召集 Swi6，Swi6 引起异染色质化。在植物和哺乳动物中，这一过程由 DNA 甲基化支持，DNA 甲基化同样可以召集异染色质化机器。哺乳动物中，单个基因也可以被 RNAi 所沉默，而 RNAi 靶向基因的调控区而不是编码区。这一沉默过程涉及 DNA 的甲基化而不是对 mRNA 的降解。

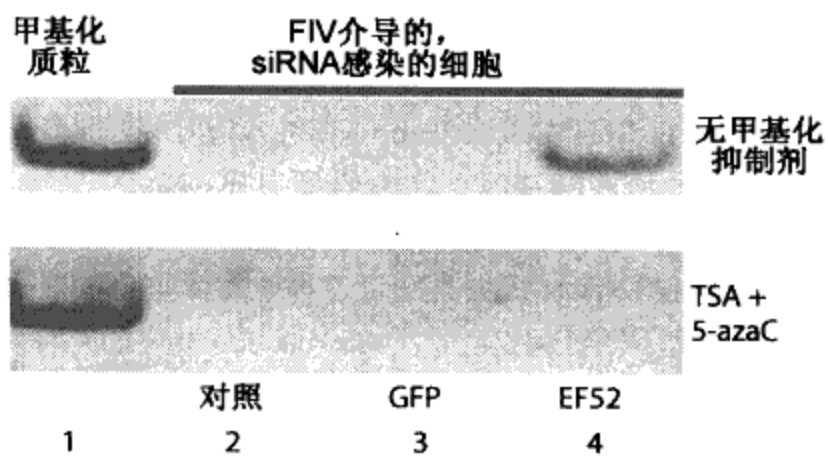


图 16.42 EF1A 基因调控区的甲基化应答 siRNA 的证明。Morris 及同事利用 *Hin*P1I 的切割特性检测 EF1A 基因调控区 CpG 序列发生甲基化的情况，*Hin*P1I 切割未甲基化的含有 CpG 序列的位点而不能切割甲基化的位点。从未经处理的细胞和经 TSA 和 5-azaC 处理的细胞中提取 DNA，TSA 和 5-azaC 可以抑制 CpG 序列甲基化。用 *Hin*P1I 处理这些 DNA，然后用 CpG 两侧序列设计引物来 PCR 扩增上述酶切产物。未切割的（甲基化的）的 DNA 才会产生 PCR 信号。泳道 1 表示含有甲基化的阳性对照。泳道 2 以不相关 RNA 为阴性对照。泳道 3 加入靶向 GFP 编码区的 siRNA 作为阴性对照。泳道 4 加入靶向 GFP 调控区的 siRNA 所产生的实验结果。在该 siRNA 存在且无 TSA 和 5-azaC 条件下，CpG 序列被甲基化（未被切割因而产生 PCR 信号）。但在甲基化抑制剂存在时，该位点未被甲基化。（*Source*：Reprinted with permission from *Science*, vol. 305, Kevin V. Morris, Simon W.-L. Chan, Steven E. Jacobsen, and David J. Looney, "Small interfering RNA-induced Transcriptional Gene Silencing in Human Cells," Fig. 1, p. 1290, Copyright 2004, AAAS.）

microRNA 和基因沉默

siRNA 不是参与基因沉默的唯一小 RNA。另一类被称为 **microRNA（miRNA）** 的小 RNA 是植物和动物细胞中自然产生的 18～25nt 的 RNA，它们是由长约 75nt 的茎环结构前体 RNA 被 Dicer RNase 酶切而成。在哺乳动物中，这些 miRNA 同特异 mRNA的 3′-UTR 碱基配对（虽然不完美），主要通过阻止这些 mRNA 的翻译来沉默基因的表达。

miRNA 的重要性最早被引起关注是始于 1981 年的研究工作，该研究表明线虫 *lin-4* 基因的突变导致线虫发育异常。随后的遗传学研究表明，*lin-4* 基

因产物通过抑制 LIN-14 的表达水平发挥作用，LIN-14 是 *lin-14* 基因的表达产物。有趣的是，*lin-4* 需要 *lin-14* 的 3′-UTR 才能发挥它对 LIN-14 的抑制功能。最后在 1993 年，Victor Ambros 等定位了 *lin-4* 突变体，发现它所在的基因不能编码蛋白质，而是一个编码 miRNA 前体的基因。这表明通过减少 LIN-14 的表达，miRNA 在线虫的发育过程中发挥重要的作用。线虫基因组序列验证了这一假说，miRNA 同 *lin-14* mRNA 的 3′-UTR 部分互补，而互补的这一序列是 *lin*4 发挥正常功能所需要的。

1999 年，Philip Olsen 等首先提出 *lin-4* 通过限制 *lin-14* mRNA 的翻译来发挥作用。LIN-14 在线虫正常发育中发挥重要作用。在第一幼虫时期（L1），LIN-14 的表达水平很高，因为该蛋白质有助于特化 L1 时期细胞的命运。然而在 L1 的末期，LIN-14 的水平下降以便其他的蛋白质在第二幼虫发育阶段（L2）来决定细胞的命运。抑制 LIN-14 的表达水平依赖于 *lin-4* RNA 的作用，*lin-4* 是一个 22nt 的 miRNA，它可以同 *lin-14* mRNA 的 3′-UTR 发生部分碱基互补配对。

Olsen 等进行了 Western 印迹实验，结果显示从 L1 发育到 L2 时期，LIN-14 的表达量至少下降了 10 倍。另外他们的核连缀转录实验显示 *lin-14* mRNA 水平仅下降了约 1/2。分析显示两时期 mRNA 的 poly（A）尾巴没有发生改变。因此，对 *lin-14* 的表达控制可能发生在翻译水平而不是转录水平。

接着，Olsen 等利用 RT-PCR 技术扩增两时期 *lin-14* mRNA 中 poly（A）的长度。结果显示两个发育时期的 poly（A）的长度没有发生改变。因此证明也不是通过剪短 poly（A）的长度来改变 L2 时期 *lin-14* mRNA 的稳定性。事实上，Olsen 的研究还显示，L2 时期同 *lin-14* 结合的多核糖核蛋白与 L1 时期一样多。因此，*lin-14* mRNA 的翻译起始应该像 L1 时期一样正常进行。

LIN-14 蛋白质在 L2 时期较少出现，而 mRNA 的翻译起始是正常的，一个合理的推论是，该 mRNA翻译的延伸和终止受到某种方式的抑制。若 *lin-4* miRNA 确实同 *lin-14* mRNA 的 3′-UTR 目标区域结合，它所处的位置刚好可以干扰 mRNA 翻译终止。这样一来，*lin-4* miRNA 和 *lin-14* mRNA 应同时出现在多聚核糖体中。

为验证这一假说，Olsen 等利用蔗糖浓度梯度超速离心纯化 L1 和 L2 的多聚核糖体，通过 RNase 保护实验（第 5 章）检测复合物是否存在 *lin-4* miRNA和 *lin-14* mRNA，结果如图 16.43 所示。顶部的曲线图中，靠近右边的驼峰状曲线含有快速沉淀的多聚核糖体。

图 16.43 下面的电泳图表示 RNase 保护实验的结果，中间的两个泳道同样含有多聚核糖体。从图 16.43 中可以看到，L1 和 L2 两发育时期出现的多聚核糖体几乎相同，同时它们几乎含有等量的 *lin-4* miRNA（中图所示）和 *lin-14* mRNA（最底部图所示），可能由于两个 RNA 互补配对所致。

这些结果出现了难以解释的问题，*lin-4* miRNA 和 *lin-14* mRNA 被发现共同存在于多聚核糖体中，表明它们碱基互补配对在一起。但 L1 和 L2 时期多聚核糖体出现的曲线图看起来是一样的。若 miRNA 完全或几乎完全阻断翻译延伸，较少的核糖体结合到 mRNA 上，多聚核糖体的会变轻，峰将会左移。然而没有观察到峰的左移现象。另一方面，若 miRNA 对翻译延伸实施了较为缓和的抑制或 miRNA 阻断了翻译的终止，则多聚核糖体会聚集较多的核糖体，使得多聚核糖体峰右移。但也没有发现峰的右移的现象。因此，*lin-4* miRNA 并不是通过简单地抑制翻译的延伸或翻译的终止来限制 L2 时期 *lin-14* 的蛋白质浓度。可以想象，*lin-4* 应该以一种不改变核糖体聚集的方式来抑制翻译的起始和延伸。还有可能是，通过与 mRNA 的 3′端结合，*lin-4* 捕获新合成的 LIN-14 蛋白并使之被降解掉。

2005 年 Amy Pasquinelli 等的报道至少解释了该问题中关于 *lin-4* miRNA 活性的部分。对线虫 RNA 进行 Northern 杂交，研究人员发现 *lin-14*（和 *lin-28*）的 mRNA 水平在 L2 中下降为原先的 1/4（图 16.44）。这些数据表明，mRNA 水平的降低依赖于 *lin-4* miRNA 存在，最多仅有少许的 mRNA 水平的下降发生在 *lin-4* e912 突变体中。因此，*lin-4* 通过不止一种机制实现它的调控功能。

尽管我们还不能精细地了解类似 *lin-4* 产物的 miRNA 是如何抑制翻译的，甚至不知道它们是否主要靠抑制翻译来发挥作用，但我们知道 miRNA 在动植物基因的表达调控中发挥着至关重要的作用。在大多数动植物中都含有上百个 miRNA，每个 miRNA 都可能控制许多其他基因的表达。miRNA 基因的突变会对生物体产生巨大的影响，尤其对生长发育，这一事实说明 miRNA 是多么的重要。至少在动物中，小 RNA 同目标基因的碱基互补程度而不是小 RNA 的起源，决定着沉默发生的类型。若碱基完全配对，则 mRNA 趋向于被降解，即使该小 RNA 是 miRNA 而不是 siRNA。若碱基不完全配对，趋向于阻断 mRNA 的翻译，即使该小 RNA 是 siRNA 而不是 miRNA。miRNA 和 mRNA 完全配对导致 mRNA 降解的好例子是小鼠中的 miR-196 和 HOXB8 mRNA。哺乳动物以及其他动物都拥有成簇的同源异型基因，它们编码具有同源域的转录因子。

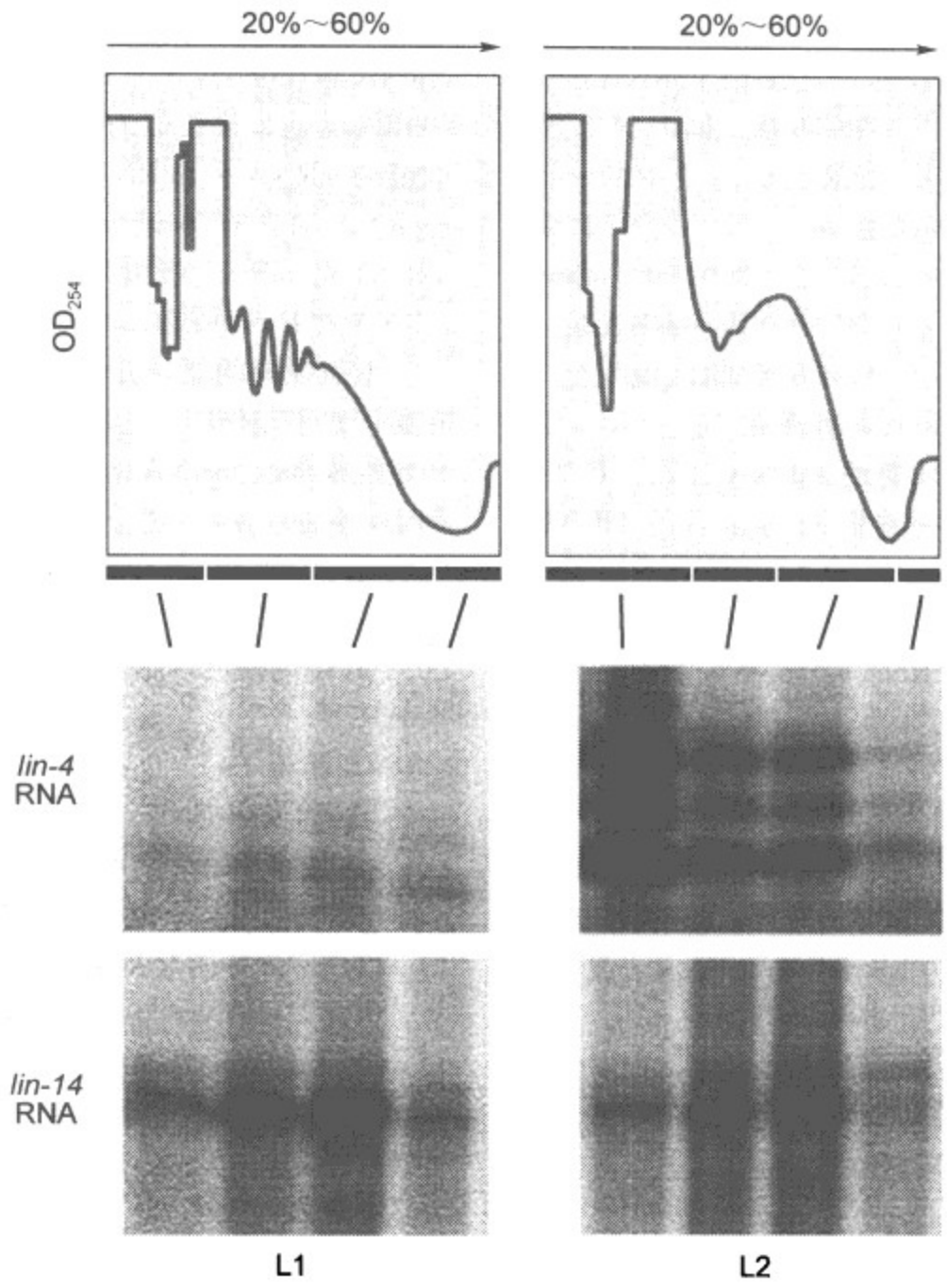
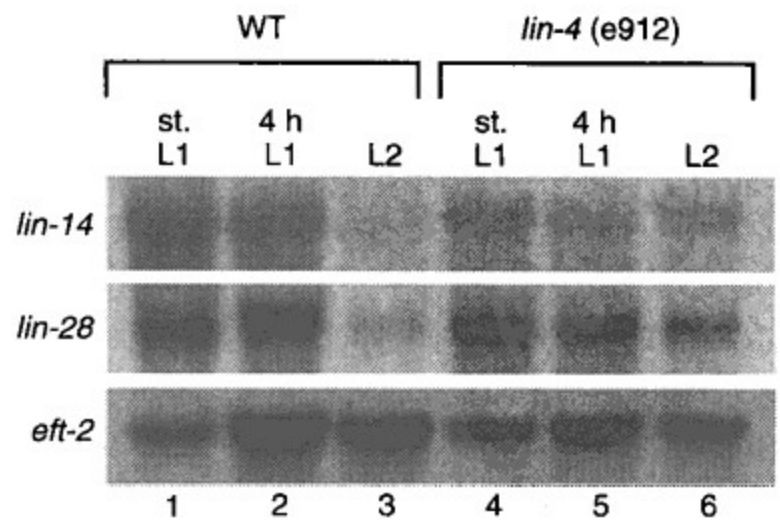


图 16.43　在 L1 和 L2 幼虫时期 *lin-4* miRNA 和 *lin-14* mRNA 都结合于多聚核糖体。Olsen 和 Ambros 用蔗糖浓度梯度超速离心的方法分离线虫 L1（左）和 L2（右）幼虫时期的多聚核糖体。收集梯度中的 4 个区段成分，中间的两个含有多聚核糖体，用这些成分中的 RNA 同来自 *lin-4* 和 *lin-14* 的 RNA 探针进行杂交。用 RNase 处理杂交后产物，PAGE 凝胶电泳被保护的探针。利用 *lin-4* 和 *lin-14* 探针所产生的结果分别表示在中图和下图中。多条带表示被保护的仅差一个核苷酸的探针，可能是由于杂交链末端的核苷酸被 RNase 一个一个地剪去所致。(*Source*: *Developmental Biology*, Volume 216, Philip H. Olsen and Victor Ambros, "The *lin-4* Regulatory RNA Controls Developmental Timing in Caenorhabditis elegans by Blocking LIN-14 Protein Synthesis after the Initiation of Translation." fig. 8, pp. 671-680, Copyright 1999, with permission from Elsevier.)

图 16.44　线虫发育时期各 mRNA 的浓度。Pasquinelli 和他的同事用来自线虫不同发育时期的 RNA 制备 Northern 杂交膜，所选时期如图顶端所示：饥饿 L1，4h L1 和 L2 时期。用 *lin-14* mRNA 和 *lin-28* mRNA 探针杂交制备的膜，同时用 *eft-2* mRNA 探针作对照，已知该 mRNA 不受 *lin-4* 的影响。野生型细胞中，*lin-14* mRNA 和 *lin-28* mRNA 的浓度在 L1 和 L2 之间的差异很大，而在 *lin-4* (e912) 细胞中差异不大。(*Source*: Reprinted from *Cell*, Vol 122, Shveta Bagga, John Bracht, Shaun Hunter, Katlin Massirer, Janette Holtz, Rachel Eachus, and Amy E. Pasquinelli, "Regulation by let-7 and *lin-4* miRNAs Results in Target mRNA Degradation," pp. 553-563, fig. 6a, Copyright 2005, with permission from Elsevier.)



这些转录因子在胚胎发育中发挥着至关重要的作用。在 HOX 基因簇中存在一些编码 miRNA 的基因，产生的 miRNA 可以下调 HOX（同源异形盒，homeobox）基因的表达。miR-196 是其中的一个 miRNA，除一个 G-U 摇摆配对外，它可以同 HOXB8 mRNA 完全互补配对。2004 年，David Bartel 等运用 RACE 技术检测 HOXB8 mRNA 的 5′端，该 mRNA 同 miR-196 互补配对的区域发生了剪切。鉴于小鼠的 miR-196 出现在 15～17d 的胚中，他们观察了这一时期的 mRNA 片段。RACE 实验获得了 8 个克隆，它们同剪切后的 HOXB8 mRNA 相对应。其中 7 个克隆的末端处于同 miR-196 碱基配对的区域。

这些结果表明，miRNA 对 mRNA 的剪切发生在两个 RNA 互补配对的区域内。为验证这一假说，Bartel 等将 miR-196 的互补序列融合到一个萤火虫萤光素酶基因中，用该基因转染 HeLa 细胞，同时分别共转染 miR-196 和无关的 miRNA。然后他们利用 RACE 技术检测报告基因 mRNA 的剪切情况。发现 miR-196 而不是无关的 miRNA，导致了萤光素酶 mRNA 的剪切。可见，若哺乳动物的 miRNA 可以和目标基因完全配对或几乎完全配对，它们就可以导致目标 mRNA 的降解。

注意 siRNA 和 miRNA 在哺乳动物中发挥作用的 3 个重要区别。

1. siRNA 通过诱导目标 mRNA 的降解来沉默基因的表达，而 miRNA 更倾向于干扰目标 mRNA 蛋白质产物的积累来沉默基因的表达。但是，如果动物 miRNA 同目标 mRNA 完全互补配对或几乎完全互补配对时，miRNA 能够导致目标 mRNA 的降解。

2. siRNA 由 Dicer 酶加工 dsRNA 所形成，该 dsRNA 中至少有一条链是外源的或来自转座子。另一方面，miRNA 由 Dicer 酶加工具有茎环结构的 RNA 的双链部分所形成，而该茎环 RNA 是正常的细胞产物。

3. siRNA 同目标 mRNA 完全互补配对，而 miRNA 往往同其目标 mRNA 不完全互补配对。

两类小 RNA（miRNA 和 siRNA）所引起的基因沉默都依赖于一个 RISC 复合体。在果蝇中存在两个 Dicer 酶（Dicer1 和 Dicer2）和两个 RISC，即 siRISC 和 miRISC，但并不是简单的一一对应关系。siRNA 引起的沉默需要 siRISC 和两个 Dicer，而 Dicer2 在产生 siRNA 方面更加重要。miRNA 引起的沉默需要 miRISC，而 miRNA 的产生仅需要 Dicer1。然而这一分工不是普遍存在的机制，因为其他生物包括酵母和哺乳动物仅仅含有一个 RISC。尽管它们的作用机制非常复杂，但越来越清楚的是，siRNA 和 miRNA 介导 mRNA 降解的基本机制即使不相同也是非常相似的。它们都需要 Dicer 酶产生双链的 siRNA 和 miRNA，这些双链 RNA 中一个单链 RNA 同含有 Argonaute 的 RISC 相结合。单链 siRNA 或 miRNA 吸引与之互补的 mRNA，继而该 mRNA 被 RISC 降解。

重点强调的是，并不是所有的动物 miRNA 都在翻译水平起作用。它们也可以通过降低 mRNA 的稳定性来降低 mRNA 的浓度。我们已经看到过两个例子，包括 *lin-4*，这个在 miRNA 家族中首先发现的成员，它既可以降低 mRNA 的浓度也可以抑制 mRNA 的翻译。在第 24 章我们将学到，用两个 miRNA 中的任何一个转染 HeLa 细胞将导致约 100 个 mRNA 水平的降低。事实上，在脑中正常表达的 miRNA 使得 HeLa 细胞的 mRNA 表达谱类似于脑中的 mRNA 表达谱。另一在肌肉中正常表达的 miRNA 使得 HeLa 细胞 mRNA 的表达谱同肌肉的 mRNA 表达谱十分相近。此外，不稳定的 mRNA 的 3′-UTR 含有的序列同各 miRNA 的 5′端序列互补。因此，miRNA 与目标 mRNA 的碱基互补配对对 mRNA 的稳定性至关重要。事实上每个 miRNA 分子可以直接或间接地影响近 100 个 mRNA 分子的表达水平，同时也表明 miRNA 在调控动物基因的表达方面发挥着十分广泛的作用，其重要性可与蛋白质转录因子相媲美。

miRNA 的发现以及其在降解 mRNA 方面的功能有利于阐明富含 AU 元件（AU-Rich Elements，ARE）的作用，早在 1986 年人们就发现 ARE 元件位于一些不稳定 mRNA 的 3′-UTR。2005 年，Jiahua Han 等报道了果蝇肿瘤坏死因子（NEF）α mRNA 的不稳定性主要受 Dicer-1、Ago1 和 Ago2 的调控，而这些蛋白质都参与 miRNA 介导的 mRNA 降解。他们继续研究表明，人类含 ARE 元件mRNA 的稳定性也受到 Dicer 的调控。此外，一个特异的 miRNA（mi-R16）同 ARE（AAUAUUUA）碱基互补，而 ARE 又是 mRNA 去稳定性所必需的。

与动物中的翻译阻断模式相比，植物中的 miRNA通过与目标基因完全配对导致这些 mRNA 的降解进而沉默目标基因的表达。例如，James Carrington 等（2002）发现一个 21nt 的 miRNA（miRNA39）在芥类植物拟南芥的开花组织中大量积累，它可以同 SCL 转录因子家族的一些成员的 mRNA 碱基互补，互补配对发生在 mRNA 的中间区段。碱基互补的结果导致 mRNA 的断裂，断裂发生在同 miRNA 互补的区域。在叶和茎组织中 miRNA39的积累较少，没有检测到 SCL mRNA 的

降解。

为阐明 miRNA 介导的 mRNA 降解，Carrington 等将编码 miRNA39 前体的基因导入叶片组织。检测到高水平的 miRNA39，说明叶组织中含有 Dicer 样酶，能将其前体加工成成熟的 miRNA。更重要的是，他们发现在表达 miRNA39 的叶组织中 SCL mRNA 被剪切成较小的无活性产物。

另一方面，一些植物 miRNA 虽然能够同目标 mRNA 进行较好地碱基互补配对，但它们主要通过抑制翻译来沉默基因的表达。Xuemei Chen 在 2004 年提供了一个很好的例证，拟南芥的 miRNA172 几乎可以同来自花序同源异型基因 APETALA2 的 mRNA 进行完全碱基互补配对，但它是通过阻断翻译来沉默目标基因的表达而不是通过 mRNA 的降解。因此，无论同目标 mRNA 碱基互补的程度如何，植物 miRNA 既可以通过 mRNA 的降解也可以通过翻译的阻断来沉默基因的表达。

图 16.45 总结了当碱基不完全互补时 miRNA 的作用方式（动物中的典型方式）和当完全配对时 miRNA 的作用方式（植物中的典型方式，动物中有时也发生）。前一种作用方式中，蛋白质的翻译至少使蛋白质产物的出现受到阻断。后一种作用方式，mRNA 被切割降解。然而，在我们的脑海中应保持一种这样的认识，即这些标准途径都有例外。动物 miRNA 虽不能与目标 mRNA 碱基完全互补但可以导致目标 mRNA 的降解，同样，植物 miRNA 虽能够同目标 mRNA 碱基完全互补但仍可以阻断目标基因翻译的进行。

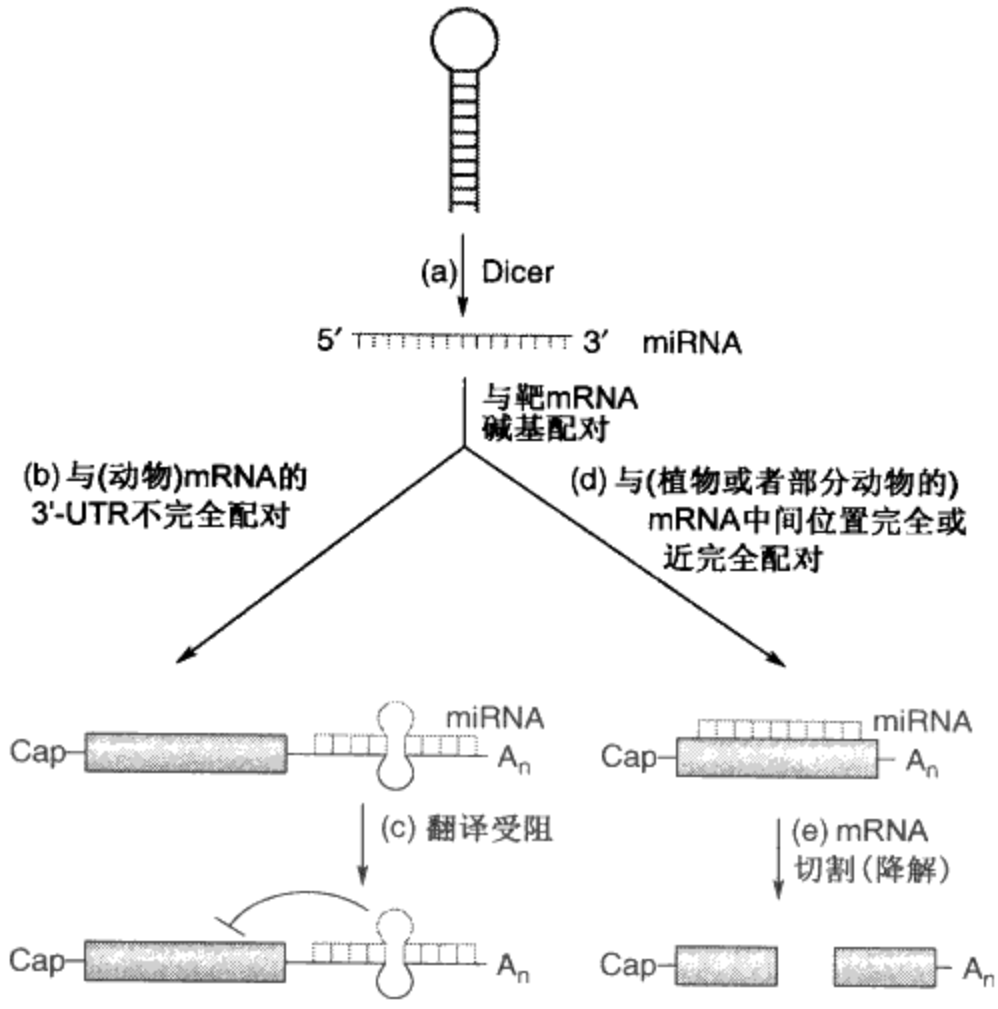


图 16.45　miRNA 介导基因沉默的两种途径。（a）茎环结构的 miRNA 前体被 Dicer 切割产生 21nt 长的短 miRNA；（b）如果 miRNA 与目标基因的 3′-UTR 不完全配对，这种情况经常发生在动物中，那么 miRNA 将导致目标基因翻译受阻，至少是影响蛋白质的积累（c）；（d）如果 miRNA 与目标基因的 3′-UTR 完全配对或几乎完全配对，这种情况常发生在植物中，动物中有时也会发生，那么 miRNA 将会引起目标 mRNA 的降解（e）。

小结　microRNA 是 18～25nt 长度的 RNA，它们来自含有茎环结构的细胞内 RNA。miRNA 形成的最后一步是，Dicer 酶切割前体 RNA 的双链茎部形成双链形式的 miRNA。其中的一条单链miRNA同 RISC 中的 Argonaute 蛋白结合，通过与目标基因 mRNA 碱基配对的方式来控制其他基因的表达。动物 miRNA 倾向于同目标mRNA 的 3′-UTR 发生不完全碱基配对，进而抑制目标 mRNA 蛋白质的积累。然而，动物 miRNA 同目标 mRNA 的完全或不完全碱基配对也可能导致 mRNA 的降解。植物 miRNA 则倾向于同目标 mRNA 发生完全或几乎完全的碱基配对，导致目标 mRNA 的裂解，尽管也存在阻断翻译的方式。

总结

核糖体 RNA 在真核细胞的细胞核中以前体的形式合成，它必须要经过进一步加工，才能产生成熟的 rRNA。尽管不同生物中成熟的 rRNA 大小不同，但是在所有的真核生物中前体 RNA 的顺序为 18S、5.8S、28S。在人类细胞中，核糖体 RNA 前体大小为 45S，它可进一步加工为 41S、32S、20S 中间产物。在加工过程中，小核仁 RNA 发挥了至关重要的作用。

tRNA 前体的 5′端多余的碱基可由内切核酸酶 RNase P 剪切掉。来自细菌和真核生物细胞核的 RNase P 有一个具有催化能力的亚基，被称为 M1 RNA 亚基。而菠菜线粒体的 RNase P 缺少这种亚基。tRNA 前体 3′端大多数多余的碱基由 RNaseII 和多聚核苷酸磷酸化酶共同催化剪切掉，只留下＋2 和＋1 处的两个碱基，它们由 RNase PH 和 RNase T 去除掉，在这期间，RNase T 起主要作用。

锥虫的 mRNA 在一个前导外显子和其他任意一个独立的外显子之间通过反式剪切的方式合成。

锥虫线粒体（动基体）编码的不完全 mRNA 在翻译前必须经过编辑。编辑通常在一系列引导 RNA 作用下沿着 3′→5′的方向进行。这些 gRNA 与待编辑的 mRNA 杂交，并提供 A 和 G 作为掺入 mRNA 中丢失 U 的模板。有时 gRNA 丢失的 A 或 G 与 mRNA 中的 U 配对的，在这种情况下，mRNA 的 U 将被切除。切除 U 的机制如下：①在切除 U 处剪切前体 mRNA；②在外切核酸酶的作用下切除 U；③将前体 mRNA 被切断的两个片段连接起来。而前体 mRNA 添加 U 的机制在第一步和最后一步与切除 U 的机制一致，只不过中间一步是在 TUTase 的催化下添加 U。

在高等真核生物中，包括果蝇和哺乳动物的 mRNA 中，一些腺苷酸为了翻译成正确的蛋白质必须要脱腺苷变成次黄嘌呤。mRNA 这种编辑是由作用于 RNA 的腺苷酸脱腺苷酶完成（adenosine deaminase active on RNA，ADAR）。另外，一些胞嘧啶为了正确编码也必须脱腺苷。

基因转录后调控的普遍方式是通过 mRNA 稳定性来调控。例如，当哺乳动物的乳腺组织受到催乳素的刺激时，酪蛋白的合成大幅度提高。然而，酪蛋白的大幅度提高不是酪蛋白基因转录速度的提高所引起的，而是由酪蛋白 mRNA 的半衰期的延长引起的。

另一个对 mRNA 稳定性研究比较透彻的例子是铁离子转运蛋白受体的 mRNA。当细胞有足够的铁离子时，铁离子转运蛋白受体数量下降以避免细胞内积累过量的铁离子。相反，当细胞缺乏铁离子时，铁离子转运蛋白受体数量上升，以转运尽可能多的铁离子进入细胞。铁离子转运蛋白受体（Tfr）基因的表达通过调控 mRNA 的稳定性来控制，其机制如下：Tfr mRNA 的 3′-UTR 区包括有 5 个被称为铁离子应答元件（IRE）的茎环结构。删除相连的两个或多个 IRE 会导致细胞丧失对铁离子的应答能力。3′-UTR 区的 IRE A 和 IRE E，以及它们中间的大环结构可以被删除而不会改变细胞对铁离子的应答能力。然而，切除所有 IRE 或者两个非 IRE 的茎环结构的任意一个都会导致 TfR mRNA 组成性稳定，从而也破坏了快速转换决定元件。因此，每一个非 IRE 的茎环和至少 IRE B～IRE D 中的一个结构都是快速转换决定元件的一部分。当铁离子浓度高时，TfR mRNA 迅速降解，而当铁离子浓度低时，TfR mRNA 缓慢降解。mRNA 的稳定性上的差异大约为 20 倍左右。测定 TfR mRNA 积累速率的差异比在铁离子浓度高或低时 TfR mRNA 合成的差异（大约为 2 倍）更加重要。没有快速转换决定元件的 TRS-3 mRNA 在稳定性上为组成型（即在铁离子浓度高或低时其稳定性不发生变化）。而不能结合 IRE-结合蛋白的 TRS-4 mRNA 无论什么铁离子浓度条件下都不稳定。TfR mRNA 降解的起始似乎是一个内切核酸酶从 mRNA 的 IRE 区域离其 3′端大约 1000bp 处的切割。这种切割反应不需要 mRNA 事先脱腺苷。

铁离子浓度高时转铁蛋白受体 TfR 的浓度就会降低，并且 TfR 浓度的降低主要归因于 TfR mRNA 稳定性的降低。对铁离子的应答主要依赖于 mRNA 的 3′-UTR，它包含 5 个称为 IRE 的茎环结构。

细胞存在 dsRNA 时会发生 RNAi，这些 dsRNA 可能来自病毒、转座子或转基因（或者实验加入的 dsRNA）。这个可以触发 RNAi 的 dsRNA 被降解成大小为 21～23nt 的片段（siRNA），由一个类似 RNase III 的 Dicer 酶完成这一过程。双链 siRNA 和 Dicer 以及 Dicer 相关蛋白 R2D2 组成一个复合体 B。复合体 B 将 siRNA 传递给 RISC 装载复合体（RLC），RLC 可能将双链 siRNA 分成两个单链 siRNA，同时把引导链转运给 RNA 诱导沉默复合体（RISC），RISC 含有称为 Argonaute2（Ago2）的蛋白质。在 Ago2 PIWI 结构域的活性位点，siRNA 的引导链同目标 mRNA 碱基互补配对，其中 PIWI 结构域是一个类似于 RNase H 的酶，也被称为 Slicer。Slicer 在目标基因同 siRNA 互补配对的区域中间切割 mRNA。在一个依赖于 ATP 的步骤中，被切割

的 mRNA 从 RISC 上释放出来，以便接受一个新的待降解 mRNA。在 RNAi 过程中，反义 siRNA 杂交到目标 mRNA 上，在依赖于 RNA 的 RNA 聚合酶的作用下，扩增得到反义 RNA 全长。这些 dsRNA 被 Dicer 消化剪切成新的 siRNA 片段，进而实现 siRNA 的扩增。

酵母着丝粒和沉默结合型区域以及其他物种中的异染色质化都涉及 RNAi 体系。在裂殖酵母着丝粒的最外面区域无义链发生转录。偶尔发生的正向转录或由 RdRP 所形成的正向转录物同反向转录物碱基配对，触发 RNAi，继而募集组蛋白甲基转移酶，组蛋白甲基转移酶使组蛋白 H3 第九位点赖氨酸发生甲基化，发生甲基化组蛋白 H3 募集 Swi6 而导致异染色质化的产生。在植物和哺乳动物中，这一过程同 DNA 的甲基化相伴而行，DNA 甲基化同样可以募集异染色质化体系。哺乳动物中的单个基因可以被 RNAi 所沉默，RNAi 的目标区域在基因调控区而不是编码区。这一沉默过程涉及 DNA 的甲基化而不是对 mRNA 的降解破坏。

microRNA 是 18～25nt 长度的 RNA，它们来自含有茎环结构的细胞内 RNA。miRNA 形成的最后一步是，Dicer 酶切割前体 RNA 的双链茎部分形成双链形式的 miRNA。其中的一条单链 miRNA 同 RISC 中的 Argonaute 蛋白结合通过与目标基因 mRNA 碱基配对的方式来控制其他基因的表达。动物 miRNA 倾向于同目标 mRNA 的 3′-UTR 发生不完全碱基配对进而抑制目标 mRNA 蛋白质的积累。然而，动物 miRNA 同目标 mRNA 的完全或不完全碱基配对也可能导致 mRNA 的降解。植物 miRNA 则倾向于同目标 mRNA 发生完全或几乎完全的碱基配对，导致目标 mRNA 的裂解，尽管也存在阻断翻译的方式。

复习题

1. 绘出哺乳动物 rRNA 前体的结构，并标注所有 3 个成熟 rRNA 的位置。

2. RNase P 的功能是什么，该酶的独特性是什么？

3. 阐述顺式剪接和反式剪接的区别。

4. 设计实验证明锥体虫的前体 RNA 剪接过程中存在 Y 形的中间体，并描述该实验的结果。试述该结果如何同反式剪接相一致而不是顺式剪接？

5. 描述 RNA 编辑的概念。什么是模糊基因或隐秘基因？

6. 设计实验证明动基体 mRNA 的编辑是按3′→5′方向进行的，试述该实验的结果。

7. 根据已掌握的知识画出 RNA 编辑的模式图。该过程涉及哪些酶？

8. 试述指导 RNA（guide RNA）存在的直接证据。

9. 举例证明鼠 GluR-B 转录本被 ADAR2 编辑是必要的，且该转录本是 ADAR2 唯一重要的目标分子。

10. 设计实验证明催乳素主要在转录后水平上调控酪蛋白基因的表达，描述该实验的结果。

11. 哺乳动物细胞中同维持铁离子稳态最直接相关的两个蛋白质是什么？它们的表达水平是如何随着铁离子的浓度改变而发生改变的？

12. 我们如何得知存在一个蛋白质同转铁蛋白受体 mRNA 的铁离子应答元件结合？

13. 请描述当转铁蛋白铁离子应答元件发生突变后，产生了不能与铁离子发生应答的稳定 mRNA 和不能与铁离子发生应答并且不稳定的 mRNA 的实验结果。并用快速翻转决定因子和 IRE 结合同 mRNA的相互作用来解释这一结果。

14. 设计实验证明铁离子应答确实依赖于的 TfR mRNA 不稳定性。

15. 提供一个模型来说明 TfR mRNA 的稳定性涉及顺乌头酸酶。

16. 什么证据表明 RNA 干扰依赖于 mRNA 的降解。

17. 提供一个模型来阐述 RNA 干扰的机制。

18. 设计实验证明 Argonaute2 具有切割（slicer）活性。

19. 在 RISC 形成过程中，R2D2 的作用是什么？如果不存在 R2D2 将会怎样？

20. 提供一个模型说明裂殖酵母中的异染色质化涉及 RNAi 机制。如何修改这一模型以描述哺乳动物中相应的情况。

21. 设计一个实验，证明 siRNA 直接作用于基因的调控区域并能导致哺乳动物基因由于涉及 DNA 甲基化酶而发生沉默。

22. 比较 siRNA 和 miRNA 产生的过程。

23. 比较哺乳动物中 siRNA 和 miRNA 典型的作用方式。

24. 哺乳动物中典型的 miRNA 同目标 mRNA 的 3′-UTR 序列如果完全或几乎完全互补配对，活性将发生怎样的改变？举例说明。

分析题

1. 为什么 dicer dsRNA 不能完全阻断 RNAi？

2. 预测下列突变体对 TfR mRNA 丰度造成的影响。无论铁离子的浓度如何变化，都导致 TfR mRNA 处于高水平还是导致 TfR mRNA 持续处于

低水平，还是对 TfR mRNA 的表达丰度没有影响？

a. 阻断顺乌头酸酶产生的突变体。

b. 抑制顺乌头酸同铁离子结合的突变体。

c. 抑制顺乌头酸同 IRE 结合的突变体。

3. 引用实验证据，论述线虫中 *lin-4* miRNA 对 *lin-14* 基因表达的影响。

翻译　赵衍鑫　校译　刘智捷　郑用琏

第 17 章　蛋白质翻译机制 I：起始

tRNA 的计算机模型 *Copyright © Tripos Associates/Peter Arnold, Inc.*

翻译是核糖体解读 mRNA 中的遗传信息并按照信息的指导合成蛋白质的过程。因此，核糖体是蛋白质合成的加工厂。tRNA 作为一个适配器（adaptor）起着同等重要的作用，它在一端结合着氨基酸，另一端与 mRNA 相互作用。

我们可以将翻译过程分为 3 个阶段：起始、延伸和终止。在起始阶段，核糖体结合 mRNA，附着到匹配 tRNA 上的第一个氨基酸也结合 mRNA。在延伸阶段，核糖体每次将一个氨基酸加到正在延伸的多肽链上。在终止阶段，核糖体释放 mRNA 及合成的多肽。原核生物和真核生物的这一总体过程类似，但有明显不同，尤其是真核生物的翻译起始格外复杂。

本章涉及原核生物和真核生物翻译的起始。由于这两个系统的术语不同，分别讨论更容易一些。因此，我们先从比较简单的原核生物开始，然后学习比较复杂的真核生物的翻译起始。

17.1　细菌中翻译的起始

在翻译起始发生之前，必须具备两个重要前提。一是产生**氨酰-tRNA**（aminoacyl-tRNA）（tRNA 上附着有相关氨基酸）。换句话说，氨基酸必须共价地结合到 tRNA 上，这一过程称作 **tRNA 负载**（tRNA charging），即 tRNA 被“负载”了氨基酸。另一个重要前提是核糖体解离为两个亚基，这是必需的，因为细胞要将起始复合物装配到小亚基上，两个亚基必须分离才能使装配成为可能。

tRNA 负载

所有 tRNA 的 3′ 端都有相同的 3 个碱基（CCA），末端的这个腺苷是负载的靶位。氨基酸通过其羧基与 tRNA 末端腺苷的 2′或 3′-OH 间的酯键附着到 tRNA 上，如图 17.1 所示。负载分两步发生（图 17.2），都需要氨酰-tRNA 合成酶催化。在第一步反应（1）中，氨基酸利用 ATP 释放的能量被活化；反应产物是氨酰-AMP，焦磷酸盐副产物只是两个末端磷酸基团（β-和 γ-磷酸基），是 AMP 形成过程中 ATP 所失去的。

（1）氨基酸＋ATP→ 氨酰-AMP ＋ 焦磷酸盐（PPi）

ATP（及其他三磷酸核苷）中磷酸基团间的化学键是高能键，断裂时释放能量，被储存在氨酰-AMP（aminoacyl-AMP）中，这就是为什么我们称之为活化的氨基酸（activated amino acid）。在加载的第二步反应中，氨酰-AMP 的能量用于将氨基酸转移到 tRNA 上，形成氨酰-tRNA（aminoacyl-tRNA）。

（2）氨酰-AMP ＋ tRNA →氨酰-tRNA ＋AMP

反应（1）和（2）的总和是：

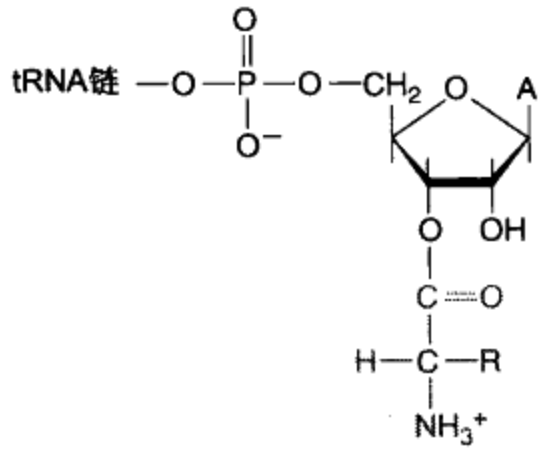


图 17.1　tRNA 和氨基酸之间的连接　有些氨基酸开始是通过酯键连接到 tRNA 末端腺苷酸的 3′-OH 上，如图所示，但有些则结合到 2′-OH 上。无论如何，氨基酸合成蛋白质之前，都要被转移到 3′-OH 上。

（3）氨基酸＋ATP ＋ tRNA→ 氨酰-tRNA ＋ AMP ＋焦磷酸盐（PPi）

与其他酶一样，氨酰-tRNA 合成酶（aminoacyl-tRNA synthetase）起双重作用，不但能催化形成氨酰-tRNA，而且还决定反应的特异性。自然界只有 20 种氨酰-tRNA 合成酶，各对应于一个氨基酸，它们非常专一，几乎总能将某个氨基酸放到正确的 tRNA 上。这种特异性对生命至关重要，如果氨酰-tRNA 合成酶出现很多错误，合成的蛋白质中将包含大量的不正确氨基酸，就不能形成正确的功能。在第 19 章将介绍该合成酶是如何正确选择 tRNA 和氨基酸的。

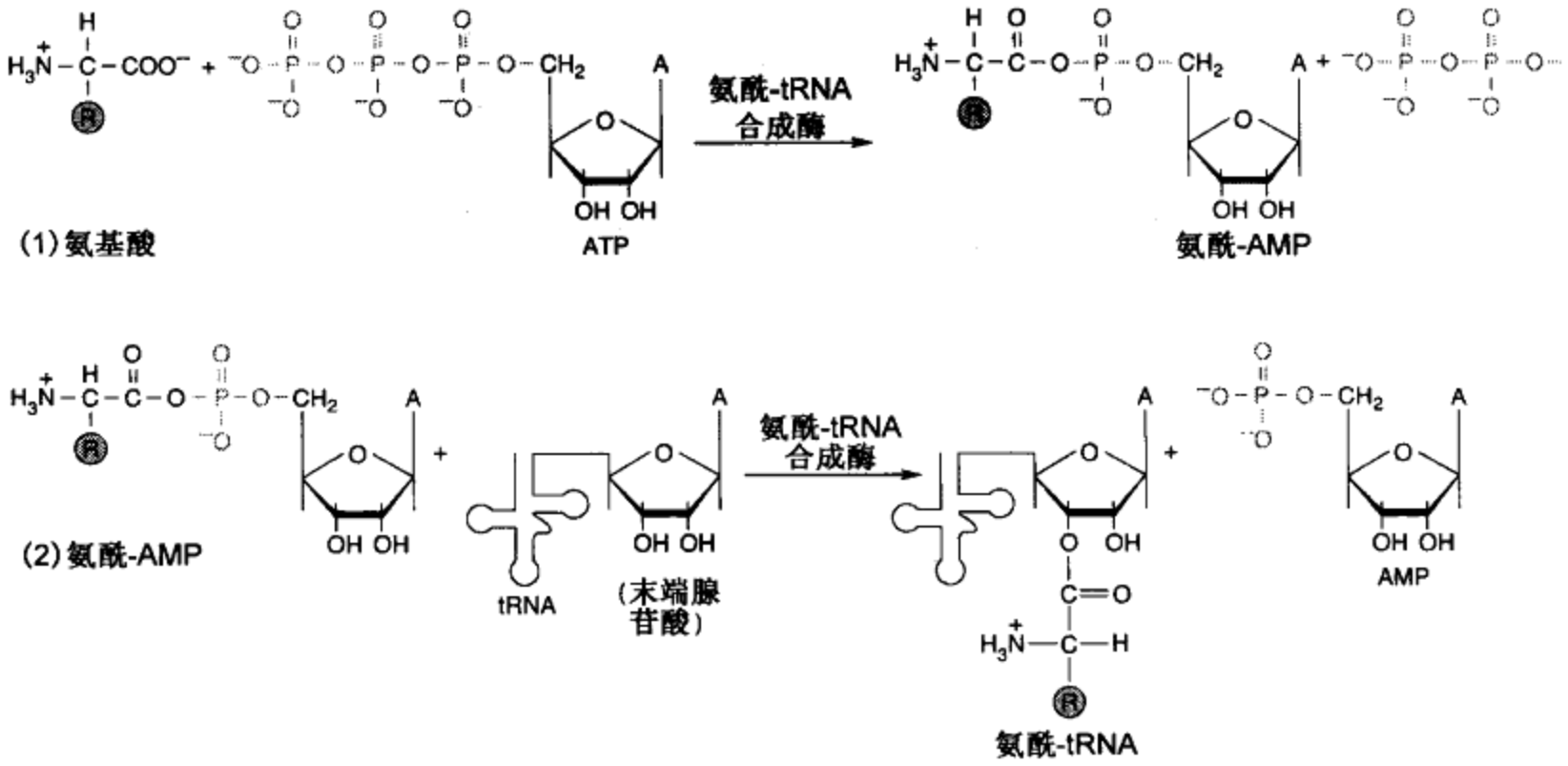


图 17.2　氨酰-tRNA 合成酶。反应（1）：氨酰-tRNA 合成酶将氨基酸连接到来自 ATP 的 AMP 上，形成氨酰-AMP，副产物是焦磷酸盐。反应（2）：合成酶用 tRNA 取代了氨酰-AMP 中的 AMP，形成氨酰-tRNA，副产物是 AMP。氨基酸加在 tRNA 末端的腺苷 3′-OH 上。

小结　氨酰-tRNA 合成酶将氨基酸连接到它们的关连 tRNA 上。这一过程通过非常特异的两步反应完成，该反应从 ATP 衍生的 AMP 活化氨基酸开始。

核糖体的解离

我们知道核糖体由两个亚基组成。例如，*E. coli*的 70S 核糖体，包含一个 30S 和一个 50S 的亚基。每个亚基有 1 或 2 个 rRNA 和大量的核糖体蛋白。30S 亚基结合 mRNA 和 tRNA 的反密码子末端，因此，它是核糖体的解码机构，阅读 mRNA 的遗传密码并与正确的氨酰-tRNA 结合。50S 亚基结合负载氨基酸的 tRNA 末端，并具有肽基转移酶活性，使氨基酸通过肽键结合到一起。

稍后我们将了解到原核与真核细胞都在核糖体小亚基上建立翻译起始复合物。这意味着对于一个新的起始复合物的形成来说，两个核糖体亚基经过每一轮的翻译后必须分离。早在 1968 年，Matthew Meselson 及同事通过图 17.3 所示的实验，提供了核糖体解离的直接证据。他们用重同位素^{15}N、^{13}C 和氘标记 *E. coli* 核糖体，加上少量的^{3}H 作为放射性示踪物。这样标记的核糖体比通常生长在^{14}N、^{12}C 和 H 培养基中的对应物更致密（图 17.4a）。接着，研究者把标记的重核糖体放在普通的氮、碳、氢培养基中培养。3.5 代以后，分离核糖体，蔗糖密度梯度离心测定其质量，以^{14}C 标记轻核糖体作为对照。图 17.4b 为实验结果。正如所预期的，他们观察到了重放射性标记的核糖体亚基（38S 和 61S 而不是标

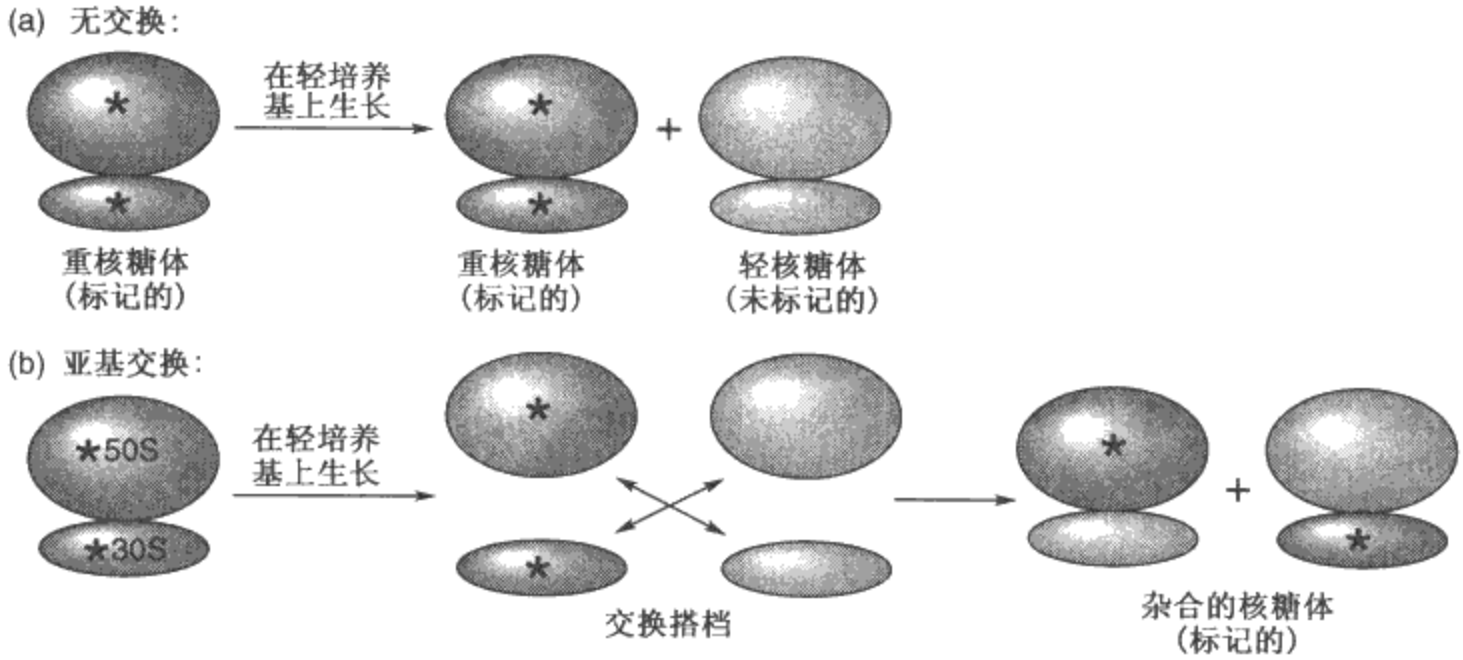


图 17.3　证明核糖体亚基交换的实验方案。Meselson 及同事在 N、C、H 的重同位素存在时培养 *E. coli* 产生重核糖体（红色），并加入少量^{3}H 使其具有放射性（星号）。然后，将带有标记的、重核糖体细胞转移到含有普通的氮、碳和氢的轻同位素培养基中。（a）没有交换。如果没有发生核糖体亚基交换，重核糖体亚基就保持在一起。观察到的唯一有标记的核糖体是重的。而在轻培养基上产生的轻核糖体因为没有放射性而检测不到。（b）亚基交换。如果核糖体解体为 50S 和 30S 亚基，重亚基能与轻亚基结合形成杂合的标记核糖体。

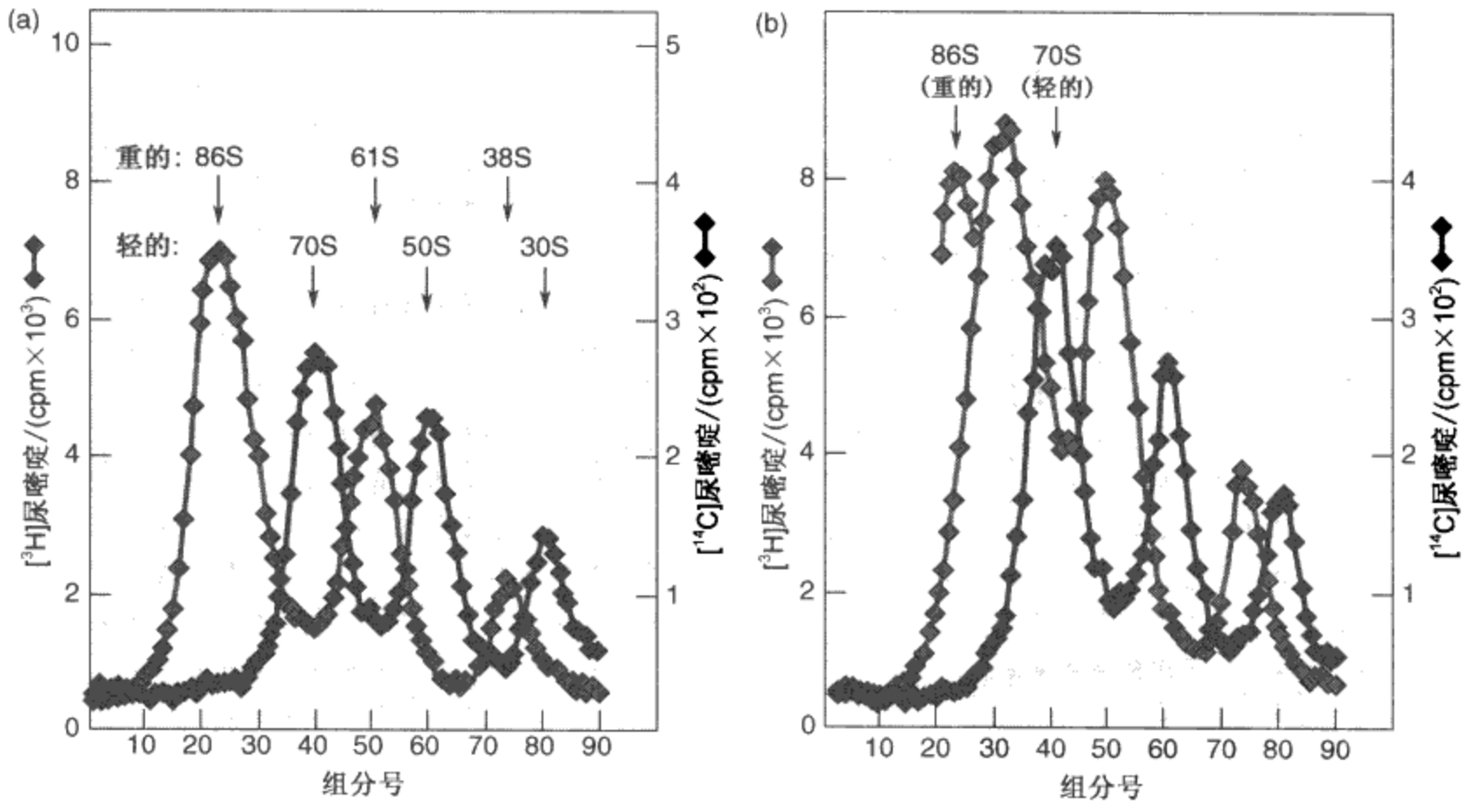


图 17.4　核糖体亚基交换的证明。（a）重和轻核糖体的沉降特点。用［^{3}H］尿嘧啶标记重核糖体（图 17.3），轻（普通）核糖体用［^{14}C］尿嘧啶标记。然后将这些核糖体进行蔗糖梯度离心，收集不同梯度的沉淀物，液体闪烁计数检测两种放射性同位素。轻核糖体和亚基（70S、50S 和 30S）与重核糖体和亚基（86S、61S 和 38S）在曲线的上方标出。（b）实验结果。培养含^{3}H 标记重核糖体的 *E. coli* 细胞，见图（a）。将这些细胞转移到普通培养基中培养 3.5 代，然后提取核糖体，加入^{14}C 标记的轻核糖体作为对照，将混合物进行蔗糖梯度超速离心，收集沉淀物并测定放射性，见图（a）：^{3}H 红色；^{14}C 蓝色。86S 重核糖体的位置以平行梯度中离心的重核糖体测定。^{3}H 标记的核糖体（最左边红色峰）是杂合的，在轻核糖体（70S）和重核糖体（86S）之间沉淀。（*Source*：Adapted from Kaempfer，R. O. R.，M. Meselson，and H. J. Raskas，Cyclic dissociation into stable subunits and reformation of ribosomes during bacterial growth，*Journal of Molecular Biology* 31：277-89，1968.）

准的 30S 和 50S)。但是标记的全核糖体的沉降系数介于标准的 70S 和两个都是重亚基的 86S 之间。这表明发生了亚基交换。重核糖体解离成亚基，又结合了新的轻亚基。

在氯化铯梯度离心上更精确地显示了有两种核糖体：一是重的大亚基加轻的小亚基，另一种是轻的大亚基加重的小亚基，如图 17.3 所预计的那样。对酵母细胞做同样实验，也得到了同样的结果。所以真核生物核糖体也在完整核糖体（80S）和亚基（40S 和 60S）之间循环。

因为生物学反应很少是自然发生的，可以预计 *E. coli* 核糖体解离成亚基需要有外界因素参与。实际上，*E. coli* 有 3 个**起始因子**（initiation factor）——**IF1**、**IF2** 和 **IF3**，其中 IF1 和 IF3 参与核糖体解离。1966 年，Severo Ochoa 及同事用盐浓度逐渐增高的缓冲液洗涤核糖体，分离到起始因子。用0.5 mol/L NaCl 分离出 IF1 和 IF2，用 1.0 mol/L NaCl 分离到 IF3。1970 年，Ochoa 及同事用蔗糖梯度离心在 IF3 或 IF1 有或无条件下测定核糖体和核糖体亚基的量，图 17.5 的结果证明高度纯化的 IF3 能明显引起*E. coli* 70S 核糖体解离为 30S 和 50S 亚基，但是 IF1 却不能，类似实验表明 IF2 也不是核糖体的解离因子。

Steven Sabol 和 Ochoa 证明 IF3 结合到 30S 核糖体颗粒上，但是当 30S 和 50S 亚基结合形成完整核糖体时它就不再结合了。用^{35}S 标记 IF3，然后以不同量加入到*E. coli* MRE600 菌株的核糖体中，该菌株在一定条件下有罕见高浓度游离的、单一的核糖体。通过蔗糖梯度离心分离核糖体和亚基。图 17.6 的结果表明，加入的 IF3 因子越多，解离发生得越多。并且标记的 IF3 专一性地与 30S 核糖体颗粒结合而与完整核糖体不结合。让结合了标记 IF3 的 30S 颗粒与 50S 颗粒结合形成核糖体，蔗糖梯度离心显示 IF3 保持附着在游离的 30S 颗粒上，完全不在完整核糖体上。因此在 30S 颗粒结合到核糖体之前，IF3 与其解离。

以上实验给我们的印象是 IF3 活跃地促进核糖体的解离，而另外两个因子没有发挥作用。但事实却不是如此：IF1 实际上促进解离，而 IF3 结合到游离的 30S 亚基上阻止其与 50S 亚基的再聚合。刚才介绍的实验没有真正地测量核糖体的解离速率，而是测量了完整核糖体与核糖体亚基的平衡浓度，这些浓度反映了解离与结合的速率。

在分析核糖体自身的解离时，Markus Noll 和 Hans Noll 发现，3 个起始因子中只有 IF1 对核糖体亚基解离速率有影响，并且这种影响是温和的，大约只有 3 倍。这些发现与图 17.7 所示过程一致：IF1 激发核糖体解离，IF3 结合到游离的 30S 亚基上阻止它们与 50S 亚基重新结合再形成核糖体。这种对亚基重新结合的干扰，对核糖体和游离亚基的平衡有根本的影响，有利于亚基的形成。

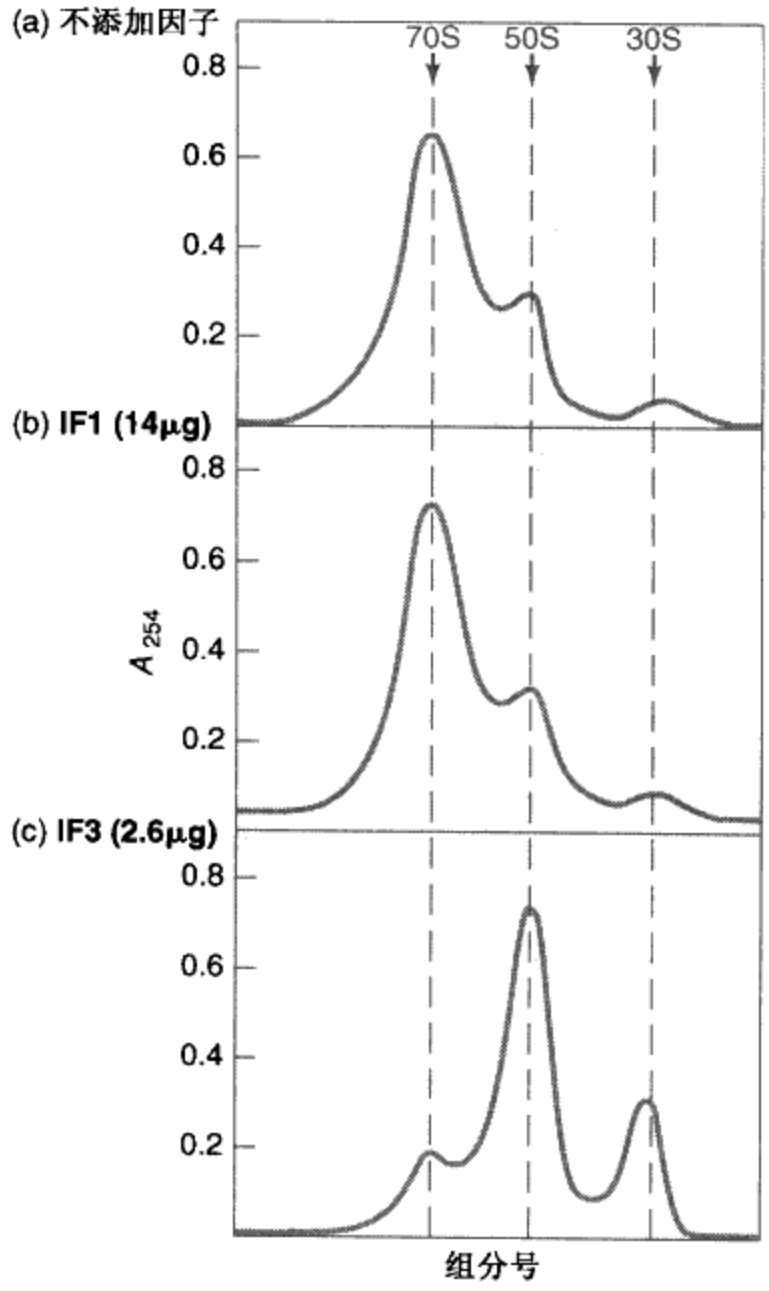


图 17.5　IF3 在核糖体解离中的作用。Ochoa 及同事将*E. coli*核糖体和下列因子混合：（a）没有因子，（b）加入 IF1，或（c）加入 IF3。然后通过蔗糖梯度超速离心从核糖体亚基中分离完整的核糖体。通过测定光吸收值（A_{254}）分析每一组沉淀中亚基的浓度，RNA（包括核糖体 RNA）在波长 254nm 时的光吸收值最大。只有 IF3 引起明显的解体。（*Source*：Adapted from Sabol，S.，M. A. G. Sillero，K. Iwasaki，and S. Ochoa，Purification and properties of initiation factor F_3. *Nature* 228：1271，1970.）

小结　在每个翻译循环的末期，*E. coli*核糖体解离为亚基。IF1 活跃地促进这种解离，IF3 结合到游离的 30S 亚基上阻止它们与 50S 亚基重新结合形成完整的核糖体。

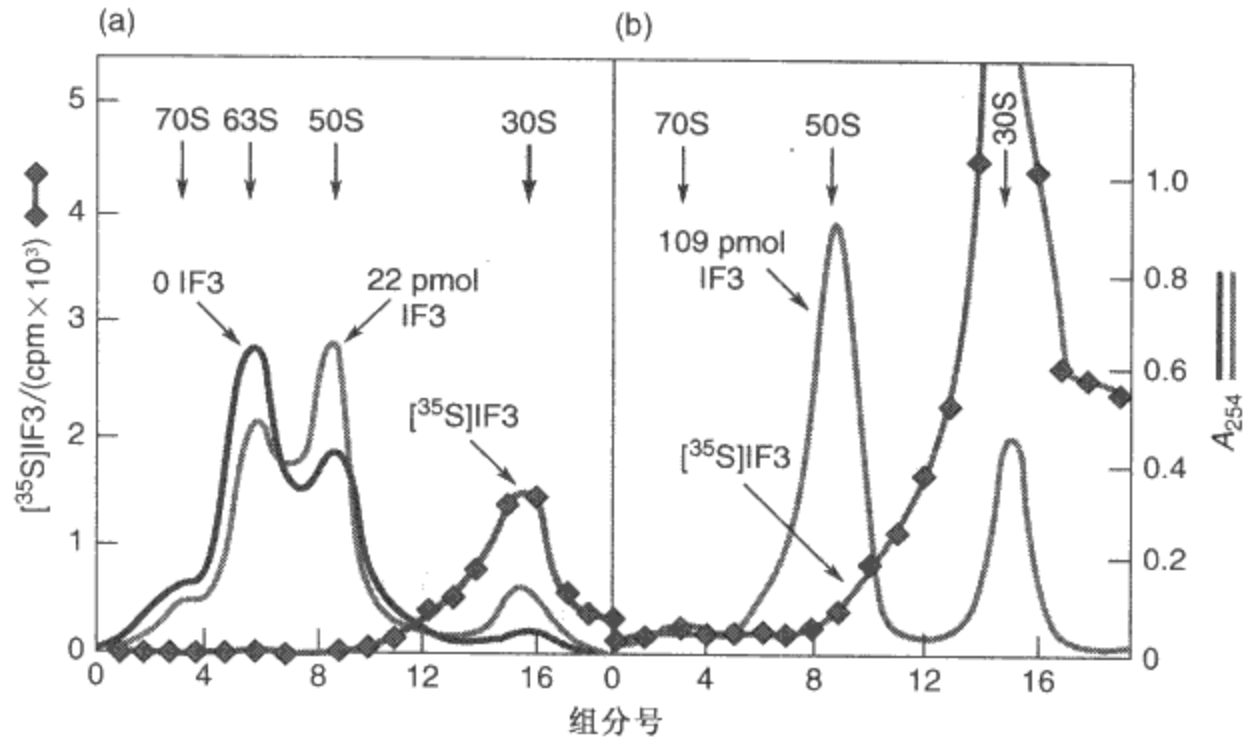
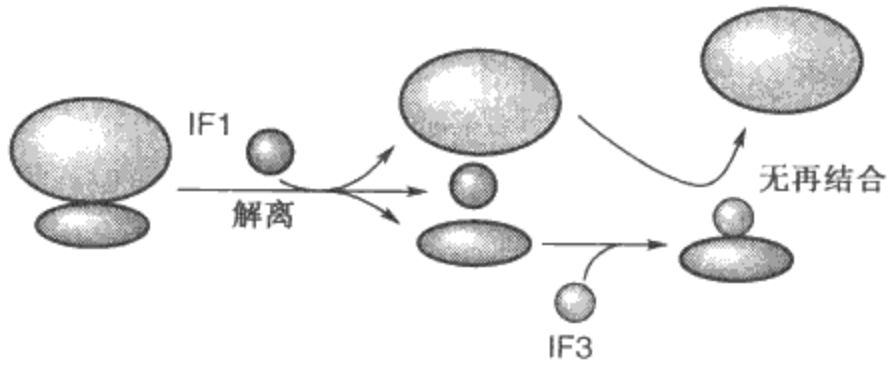


图 17.6　IF3 结合到 30S 核糖体亚基。Sabol 和 Ochoa 将*E. coli* MRE600 核糖体与不同浓度的 IF3（^{35}S 标记）混合，然后将核糖体进行蔗糖梯度超速离心，测量每一部分核糖体和亚基（A_{254}）及 IF3（^{35}S 放射能）的浓度。(a) 没有加入 IF3 的核糖体和亚基（蓝色），加入 22 pmol IF3（绿色），^{35}S 标记 IF3（红色）。(b) 加入 109 pmol IF3 的核糖体和亚基（绿色），IF3（红色）。IF3 只与 30S 亚基结合。（*Source*: Adapted from Sabol, S. and S. Ochoa, Ribosomal binding of labeled initiation factor F_3. *Nature New Biology* 234: 234, 1971.）

图 17.7　IF1 和 IF3 在核糖体解体中的作用。IF1（绿色）促进核糖体亚基（粉红和蓝色）的解体。然后 IF3（棕褐色）结合到游离的 30S 亚基上，防止 30S 亚基与游离的 50S 亚基的再结合。



30S 起始复合物的形成

一旦 IF1 和 IF3 将核糖体解离为 50S 和 30S 亚基后，细胞就在 30S 亚基上建立一个复合物，包括 mRNA、氨酰-tRNA 和几个起始因子，这就是 **30S 起始复合物**（30S initiation complex）。已经知道 IF3 能够自己结合到 30S 亚基上，而 IF1 和 IF2 稳定这种结合。同样地，IF2 能够结合到 30S 颗粒上，但在 IF1 和 IF3 的帮助下可实现更稳定的结合。IF1 自己不能结合到 30S 颗粒上，但在另两个因子协助下可以与之结合。所有 3 个因子紧密地结合在 30S 亚基靠近 16S rRNA 3′端的一个位点上。一旦 3 个起始因子结合在一起，它们就将另外两个关键的参与者 mRNA 和初始氨酰-tRNA 吸引到复合体上，后两者的结合顺序是随机的。

起始密码子和初始氨酰-tRNA　1964 年，Fritz Lipmann 证明用核糖核酸酶消化亮氨酰-tRNA 生成亮氨酸腺苷酯（图 17.8a）。这正是我们所预期的，因为氨基酸是结合在 tRNA 3′端腺苷的羟基上的。但是当 K. A. Marcker 和 Frederick Sanger 用甲硫氨酰-tRNA 做同样实验时，他们不仅发现了预期的腺苷-甲硫氨酸酯，还发现了腺苷-*N*-甲酰甲硫氨酸酯（图 17.8b）。这就证明他们实验中的 tRNA 酯酰化不仅有甲硫氨酸，而且还有甲硫氨酸的衍生物 ***N*-甲酰甲硫氨酸**（*N*-formyl-methionin，fMet）的参与。图 17.8c 比较了甲硫氨酸和 fMet 的结构。

接着，B. F. C. Clark 和 Marcker 证明*E. coli*细胞有两种不同的可负载甲硫氨酸的 tRNA。他们采用称作反向流分配的传统制备方法分离到这两种 tRNA。移动快的 tRNA，现在称为 $tRNA_m^{Met}$，可负载甲硫氨酸，但这个甲硫氨酸不能被甲酰化。就是说，其氨基不能接受甲酰基。移动慢的 tRNA 是 $tRNA_f^{Met}$，表示所结合的甲硫氨酸能被甲酰化。注意，甲硫氨酸的甲酰化发生在 tRNA 上，tRNA 不能直接负载甲酰甲硫氨酸。Clark 和 Marcker 继续做实验来验证这两种 tRNA 的两个性质：①它们对应的

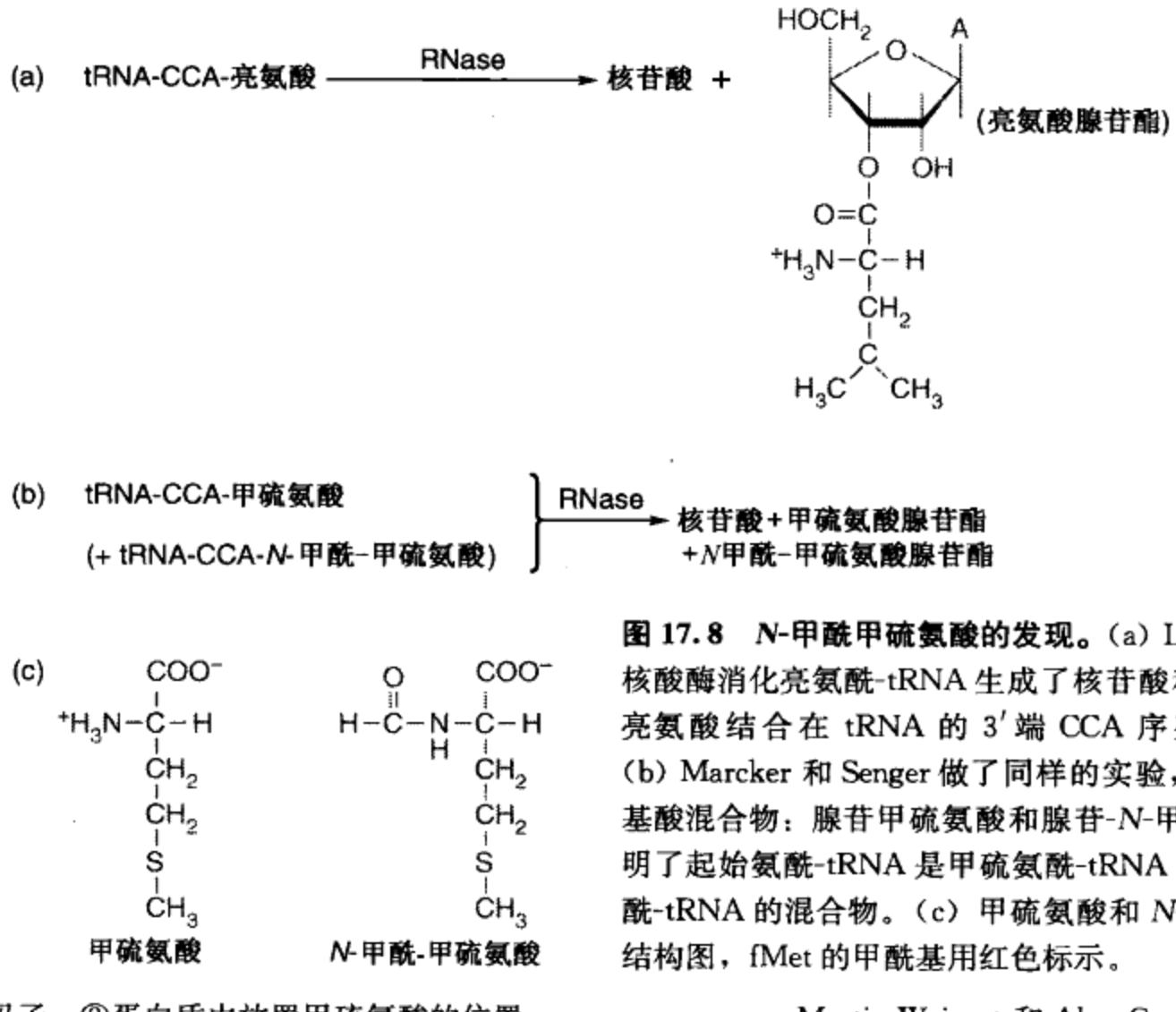


图 17.8 *N*-甲酰甲硫氨酸的发现。(a) Lipmann 等用核糖核酸酶消化亮氨酰-tRNA 生成了核苷酸和亮氨酸腺苷酯。亮氨酸结合在 tRNA 的 3′端 CCA 序列的 A 末端上。(b) Marcker 和 Senger 做了同样的实验，获得了腺苷-氨基酸混合物：腺苷甲硫氨酸和腺苷-*N*-甲酰甲硫氨酸，证明了起始氨酰-tRNA 是甲硫氨酰-tRNA 和 *N*-甲酰甲硫氨酰-tRNA 的混合物。(c) 甲硫氨酸和 *N*-甲酰甲硫氨酸的结构图，fMet 的甲酰基用红色标示。

密码子；②蛋白质内放置甲硫氨酸的位置。

密码子特异性分析采用 Marshall Nirenberg 的方法，将在下一章详述。其策略是标记氨酰-tRNA，使之与核糖体及多种三核苷酸混合，如 AUG。编码某个特定氨基酸的三核苷酸通常能使正确的氨酰-tRNA 结合到核糖体上。本例中，$tRNA_m^{Met}$ 对应密码子 AUG，而 $tRNA_f^{Met}$ 对应 AUG、GUG 和 UUG。因已知 $tRNA_f^{Met}$ 参与起始，这就表明所有这 3 个密码子 AUG、GUG 和 UUG 都能作为起始密码子。大量*E. coli*基因序列分析证实 90%以上基因的起始密码子是 AUG，8% 和 1%的基因分别以 GUG 和 UUG 为起始密码子。

随后，Clark 和 Marcker 确定了两种 tRNA 将甲硫氨酸加到蛋白质链上的位置。他们采用了体外翻译系统，该系统含有一个人工合成的 mRNA，有多个 AUG 密码子分散其中。当他们用 $tRNA_m^{Met}$ 时，甲硫氨酸主要出现在蛋白质产物的内部，用 $tRNA_f^{Met}$ 时，甲硫氨酸只进入多肽链的第一个位置。因此，$tRNA_f^{Met}$ 看起来是作为起始氨酰-tRNA 的。那么，这是由于氨基酸的甲酰化作用，还是 tRNA 的某些特性引起的呢？为了找到答案，Clark 和 Marcker 用甲酰化和未甲酰化的甲硫氨酰-$tRNA_f^{Met}$ 做进一步研究，发现是否甲酰化没有区别。两种情况下，$tRNA_f^{Met}$ 都指导了第一个氨基酸的结合。因此是甲酰甲硫氨酸-$tRNA_f^{Met}$ 中的 tRNA 部分使其成为起始氨酰-tRNA。

Martin Weigert 和 Alan Garen 用体内实验进一步证明 $tRNA_f^{Met}$ 是起始氨酰-tRNA。他们用 R17 噬菌体感染*E. coli*，分离出新合成的噬菌体外壳蛋白，发现 fMet 位于 N 端，如果它是起始氨基酸，这儿正应该是其出现的位置。丙氨酸是新外壳蛋白的第二个氨基酸，但成熟 R17 噬菌体外壳蛋白的 N 端有丙氨酸，因此，蛋白质的成熟必须包括 N 端 fMet 的切除。对多种不同细菌和噬菌体的蛋白质研究表明 fMet 是经常被切除的。有些情况下会保留甲硫氨酸，但甲酰基团总是被切除的。

小结　原核生物的起始密码子通常是 AUG，但也可以是 GUG，极少数是 UUG。起始氨酰-tRNA 是 *N*-甲酰甲硫氨酸-$tRNA_f^{Met}$。因此，*N*-甲酰甲硫氨酸（fMet）是加入多肽链的第一个氨基酸，但是，在蛋白质加工成熟过程中，fMet 是经常被切除的。

mRNA 与 30S 核糖体亚基的结合　我们已经知道起始密码子是 AUG，有时是 GUG 或 UUG，但是这些密码子也出现在遗传信息的内部。内部 AUG 编码普通的甲硫氨酸，GUG 和 UUG 分别编码缬氨酸和亮氨酸。那么，细胞怎样区别具有相同序列的起始密码子与普通密码子呢？对此问题很容易想到两种解释：在起始密码子附近要么有一个特殊的初级结构（RNA 序列），要么有一个特殊的二级 RNA 结构来判断密码子，并允许核糖体在此结合。1969

年，Joan Steitz 研究了大肠杆菌噬菌体 R17mRNA 的结构特征。噬菌体 R17 属于一类小球状 RNA 噬菌体，这类噬菌体还有 f2 和 MS2，它们是**正链噬菌体**（positive strand phage），意味着其基因组就是其 mRNA。因此，这些噬菌体提供了纯 mRNA 的方便来源，它们非常简单，只有 3 个基因，分别编码 A 蛋白（或成熟蛋白）、外壳蛋白和复制酶。Steitz 研究了 R17 mRNA 的 3 个起始密码子附近的序列，以寻找明显的初级或二级结构。在使核糖体保持在起始位点的条件下，将核糖体结合到 R17 mRNA 上，然后用 RNase A 消化没有被核糖体保护的 RNA。最后对核糖体保护的起始区域进行序列分析，发现在起始位点附近没有明显的序列或二级结构相似性。

事实上，随后的研究表明二级结构对所有 3 个起始位点都有抑制效应，松弛这些二级结构可增强起始。这在复制酶基因上尤其如此，如图 17.9a 中所示的 MS2 噬菌体 RNA，其起始密码子隐藏在外壳基因的一个双链结构内。这解释了为什么其复制酶基因的翻译是在外壳基因翻译之后进行的，即核糖体经过外壳基因时，打开了隐藏复制酶基因起始密码子的二级结构（图 17.9b）。

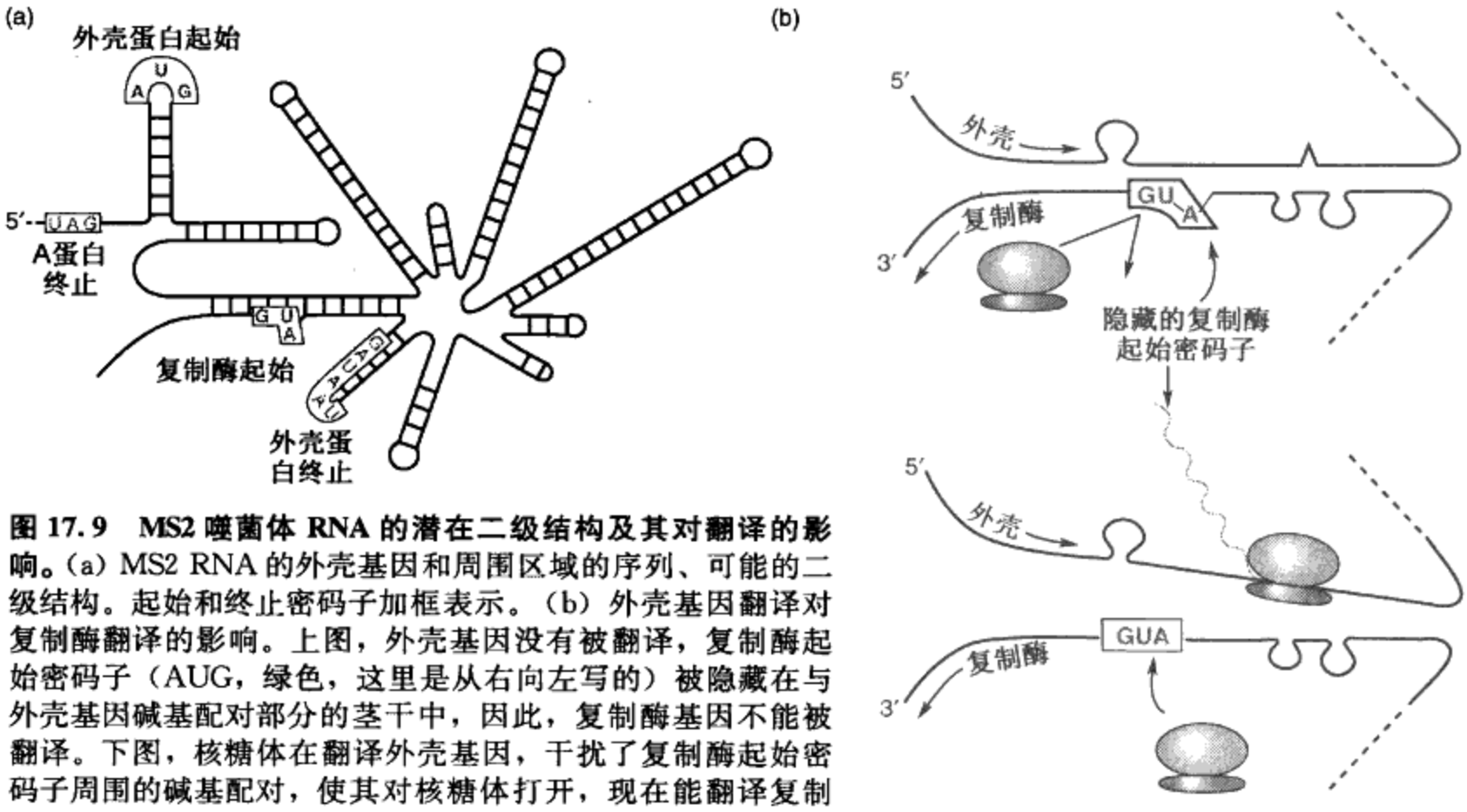


图 17.9　MS2 噬菌体 RNA 的潜在二级结构及其对翻译的影响。（a）MS2 RNA 的外壳基因和周围区域的序列、可能的二级结构。起始和终止密码子加框表示。（b）外壳基因翻译对复制酶翻译的影响。上图，外壳基因没有被翻译，复制酶起始密码子（AUG，绿色，这里是从右向左写的）被隐藏在与外壳基因碱基配对部分的茎干中，因此，复制酶基因不能被翻译。下图，核糖体在翻译外壳基因，干扰了复制酶起始密码子周围的碱基配对，使其对核糖体打开，现在能翻译复制酶基因了。（*Source*：（a）Adapted from Min Jou, W., G. Haegeman, M. Ysebaert, and W. Fiers, Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. *Nature* 237：84，1972.）

我们已经知道不能依靠二级结构来识别起始密码子，而且第一个起始位点附近的序列也没有任何明显的相似性，那么，是什么构成了核糖体的结合位点呢？答案是一个特殊序列，但有些基因（如 R17）的外壳蛋白基因偏离保守序列太多，难以识别。Richard Lodish 及同事用不同细菌的核糖体来翻译 f2 外壳 mRNA，为发现这一序列奠定了基础。他们发现，*E. coli*核糖体能够在体外翻译全部 3 个 f2 基因，而嗜热脂肪芽孢杆菌 *B. stearothermophilus* 的核糖体只能翻译 A 蛋白基因。真正的问题是外壳基因的翻译，我们知道，复制酶基因的翻译依赖于外壳基因的翻译，所以嗜热脂肪芽孢杆菌核糖体不能翻译 f2 复制酶基因只是它不能翻译外壳基因的间接效应。Lodish 及同事用核糖体组合实验证明是嗜热脂肪芽孢杆菌的核糖体，而不是其起始因子有问题。

之后，Nomura 及同事用 R17 噬菌体 RNA 进行了更细致的核糖体组合实验。发现重要元件位于 30S 核糖体亚基上。如果 30S 亚基来自*E. coli*，R17 外壳基因就能被翻译，如果 30S 亚基来自嗜热脂肪芽孢杆菌，则该基因不能被翻译。最后，他们把 30S 亚基解离成各个 RNA 和蛋白质组分，再进行组合实验。这一次，有两个成分突现出来：一是核糖体蛋白之一的 S12，一是 16S rRNA。如果两种成分中的任一个来自*E. coli*，外壳基因的翻译就处于活化状态，如果其中任一个来自嗜热脂肪芽孢杆菌，外壳基因的翻译活性就被抑制（虽然没有像整个核糖体亚基都来自嗜热脂肪芽孢杆菌那么低）。

这些发现激励了 John Shine 和 Lynn Dalgarno 去寻找 16S rRNA 与 R17 基因起始位点周围的序列间可能存在的相互作用。他们注意到所有包含全部或部分 AGGAGGU 序列的结合位点刚好位于起始密码子上游，该序列与*E. coli* 16S rRNA 3′端序列 3′-HO-AUUCCUCCAC5′的画线部分互补。注意羟基表示 16S rRNA 的 3′端，这个序列是按 3′→5′写的，所以它与 **AGGAGGU** 序列明显互补。这种关系令人感兴趣，尤其是考虑到外壳蛋白序列与 16S rRNA 的互补是在 3 个基因中最弱的，因此，很可能对 16S rRNA 序列的改变最敏感。

比较*E. coli*和 *B. stearothermophilus* 16S rRNA 的序列，发现 R17 外壳核糖体结合位点与 *B. stearothermophilus* 16S rRNA 之间匹配更差。这一结果更令人感兴趣，表 17.1 是比较分析的结果以及核糖体组合实验的分析结果，该分析是对未折叠的 R17 RNA 进行的，其二级结构不发挥作用。注意芽孢杆菌 16S rRNA 与 A 蛋白及复制酶核糖体结合位点形成 4 个 Watson-Crick 碱基对，但与外壳蛋白基因只有两个这样的碱基对。而*E. coli* 16S rRNA 与所有 3 个基因的核糖体结合位点都至少形成 3 个碱基对。那么，16S rRNA 和翻译起始位点上游区域的碱基配对对于核糖体的结合非常重要吗？如果是，就能解释 *B. stearothermophilus* 核糖体对 R17 外壳蛋白起始位点不能很好结合的原因，并且也能将 AGGAGGU 确定为核糖体结合位点。我们将看到，其他证据显示这的确是核糖体结合位点，并被称为 Shine-Dalgarno 序列或 SD 序列，以纪念其发现者。

为支持这一假说，Shine 和 Dalgarno 从另外两种细菌，绿脓杆菌（*Pseudomonas aeruginosa*）和新月柄杆菌（*Caulobacter crescentus*）（译者注：原文为 *Caulubacter crescentus*）中分离出核糖体，对 16S rRNA 3′端进行序列分析，检测了核糖体结合到 3 个 R17 起始位点的能力。与他们的其他结果一致，他们发现无论何时只要 16S rRNA 与起始密码子上游序列间有 3 个或更多的碱基配对，核糖体的结合就会发生，无论何时，少于 3 个碱基配对，就没有核糖体的结合。

Steitz 和 Karen Jakes 提供了更多有力的证据支持 Shine-Dalgarno 假说。他们将*E. coli*的核糖体结合到 R17 A 蛋白基因的起始区，然后用序列特异性的 RNase colicin E3 处理复合物，该酶切割*E. coli* 16S rRNA 的 3′端附近。接着进行 RNA 指纹实验，发现有一段双链 RNA 片段（图 17.10）。其中一条是 A 蛋白基因起始位点的寡核苷酸，包括 Shine-Dalgarno 序列，与其碱基配对的是 16S rRNA 3′端的寡核苷酸。这直接证明了 Shine-Dalgarno 序列与 16S rRNA 3′端的碱基配对，无可置疑这就是真正的核糖体结合位点了！记住，原核生物的 mRNA 通常都是多顺反子，即它们携带一个以上单顺反子或基因的信息。mRNA 中的每个顺反子都有自己的起始密码子和核糖体结合位点。因此，核糖体独立地结合到每个起始位点，并通过使某些起始位点更有利于吸引核糖体而提供了一种调控基因表达的方法。

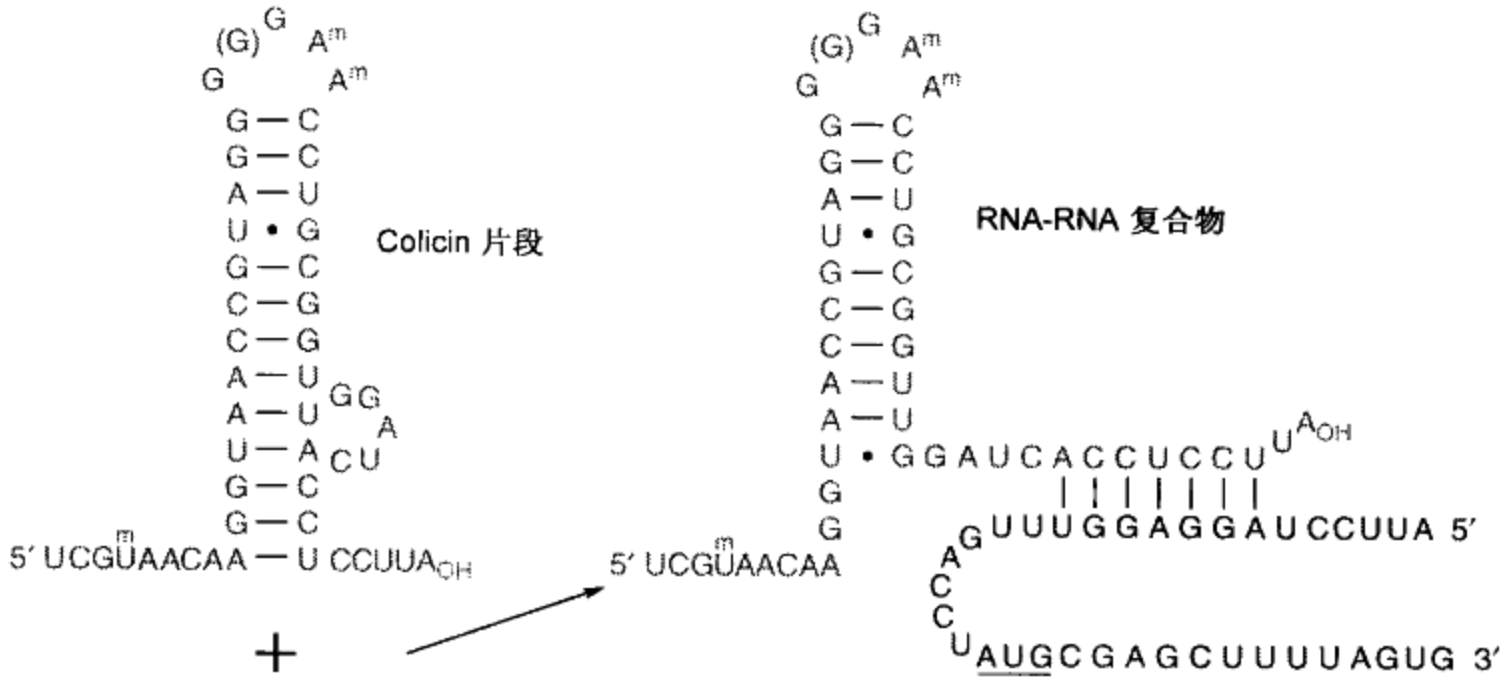


图 17.10　*E. coli* 16S rRNA 3′端 colicin 片段的可能结构和 R17 噬菌体 A 蛋白顺反子的起始区域。起始密码子（AUG）用下画线标出，colicin 片段中的“m”表示甲基化碱基。碱基间的短线表示碱基对的相对强弱：G-C>A-U>G-U。（*Source*：Adapted from Steitz, J. A. and K. Jakes, How ribosomes select initiator regions in mRNA, *Proceedings of the National Academy of Sciences USA* 72 (12): 4734-38, December 1975.）

1987年，Anna Hui 和 Herman De Boer 为 Shine-Dalgarno 序列和 16S rRNA 3′端的碱基配对提供了极好的证据。他们将突变的人生长激素基因克隆到 *E. coli*表达载体上，该载体带有一个野生型 SD 序列（GGAGG），与野生型 16S rRNA 反义 SD 序列（CCUCC）互补。这样能产生高水平的人生长激素。将 SD 序列突变为 CCUCC 或 GUGUG，它们不能与 16S rRNA 反义 SD 序列配对，两个突变体都不能产生大量人生长激素。但是当把 16S rRNA 基因的反义 SD 序列突变为 GGAGG 或 CACAC 时，又分别恢复了与 CCUCC 和 GUGUG 的互补配对。现在，带有突变体 CCUCC SD 序列的 mRNA 能被带有 GGAGG 反义 SD 序列的 16S rRNA 突变体细胞很好地翻译；带有突变 GUGUG SD 序列的 mRNA 能被带有 CACAC 反义 SD 序列的 16S rRNA 细胞很好地翻译。这种基因间的抑制很难解释，除非在这些序列间发生了碱基配对。

什么因子参与 mRNA 与 30S 亚基的结合呢？1969年，Albert Wahba 及同事证明所有 3 个起始因子都是最佳结合所必需的，但 IF3 是最重要的。他们将两种大肠杆菌噬菌体 R17 和 MS2 的^{32}P 标记的 mRNA 以及烟草花叶病毒（TMV）的 mRNA 与核糖体亚基和起始因子单独或组合在一起混合。这些病毒的 RNA 基因组都可作为 mRNA，所以是这类实验很方便的 mRNA 来源。表 17.1 的实验 1 汇总了实验结果。IF2 或 IF2＋IF1 促进 R17 mRNA 与核糖体结合的能力非常弱，但是 IF3 自身就能引起显著的结合。IF1 可以进一步刺激这一结合，3 个因子共同作用效果最好。因此，IF3 可能是 mRNA 与核糖体结合的主要因子，另外两个因子协助这个任务。我们知道 IF3 结合到 30S 亚基，使 50S 亚基不能与游离的 30S 颗粒结合，另两个起始因子也结合在 IF3 结合位点的附近，参与 30S 起始复合物的组装。

表 17.1　起始因子在 30S 起始复合物与天然 mRNA 形成中的作用

实验	核糖体	mRNA	加入的起始因子	核糖体结合/pmol	
				mRNA	fMet-tRNA$_f^{Met}$
1	30S＋50S	R17	IF1＋IF2	0.4	0.4
			IF2	0.3	0.3
			IF3	2.7	0.1
			IF1＋IF3	4.8	0.2
			IF2＋IF3	2.5	1.3
			IF1＋IF2＋IF3	6.2	6.6
2	30S	MS2	IF1＋IF3		0.0
			IF2		1.8
			IF1＋IF2		3.7
			IF2＋IF3		2.7
			IF1＋IF2＋IF3		7.3
3	30S＋50S	TMV	IF1＋IF3		0.5
			IF2		1.7
			IF1＋IF2		3.1
			IF2＋IF3		8.3
			IF1＋IF2＋IF3		16.9

Source: Role of Initiation Factors in Formation of the 30S Initiation Complex with Natural mRAN from A. J. Wahba, K. Iwasaki, M. J. Miller, S. Sabol, M. A. G. Sillero, & C. Vasquez, "Initiation of Protein Synthesis in Escherichia Coli II," *Cold Spring Harbor Symposia in Quantitative Biology*, 34:292. Copyright © 1969, Cold Spring Harbor Laboratory Press. Reprinted with permission.

小结　30S起始复合物是由游离的30S核糖体亚基加 mRNA 和 fMet-tRNA$_f^{Met}$组成的。原核细胞30S核糖体亚基与 mRNA 起始位点的结合依赖于紧接起始密码子上游被称作 Shine-Dalgarno 序列的短 RNA 序列与 16S rRNA 3′端互补序列的碱基配对。这种结合在 IF1 和 IF2 的协助下由 IF3 介导，到此，所有 3 个起始因子都已经结合到 30S 亚基上了。

fMet-tRNA$_f^{Met}$对 30S 起始复合物的结合　如果 IF3 对于 mRNA 结合到 30S 核糖体上起主要作用，那么哪个起始因子对 fMet-tRNA$_f^{Met}$起同样的作用呢？由表 17.1 可知是 IF2。IF1 和 IF3 合在一起几乎不能引起 fMet-tRNA$_f^{Met}$的结合，而 IF2 自己就能引起显著的结合。同样，如同与 mRNA 的结合一样，所有 3 个因子共同作用的效果最佳。

1971 年，Sigrid 和 Robert 证明伴随每 1mol fMet-tRNA$_f^{Met}$的结合有 1mol GTP 结合到 30S 核糖体亚基上，但是 GTP 水解是在 50S 核糖体亚基结合到复合物上且 IF2 解离之后。实验用了 3 种不同的放射性标记：^{3}H 标记 GTP 的鸟苷部分，^{32}p 标记 γ-磷酸盐，^{14}C 标记 fMet-tRNA$_f^{Met}$。将标记成分与 30S 核糖体亚基及其他因子混合，过滤混合物获得核糖体，让未结合的核苷酸和 RNA 流过。然后，检测附着在滤膜上的每种放射性同位素的量，结果见表 17.2。实验 1 证明有相同摩尔量的 GTP 和 fMet-tRNA$_f^{Met}$结合到 30S 颗粒上，IF2 对于两种结合都是必需的。实验 2 表明 GTP 对 30S 颗粒的结合依赖于三联密码 AUG 和 fMet-tRNA$_f^{Met}$，AUG 充当了 mRNA 的替代物。实验 3 确证了实验 1 和 2 的发现，同时阴性对照证实 30S 核糖体亚基是必需的。

表 17.2　GTP 对 30S 起始复合物的结合

实验	条件	结合的 pmol		
		[^{3}H]GTP	[^{32}P]GTP	[^{14}C]fMet-tRNA
1	完整	2.4		2.2
	减去 IF2	0.7		0.2
2	完整	2.1	2.0	2.3
	减去 AUG	0.6	0.2	0.1
	减去 fMet-tRNA	0.7	0.8	0.0
3	完整	2.3	2.4	2.5
	减去 IF2	0.3	0.5	0.5
	完整，加上 50S 核糖体	0.6	0.6	2.5
	减去 IF2，加上 50S 核糖体	0.3	0.5	0.4

Source: Adapted from Thach, S. S. and R. E. Thach. 1971. One molecule of guanosine triphosphate is present in each 30S initiation complex *Nature new Biology* 229: 219-20.

在实验 2 中，^{3}H 和^{32}P 标记的 GTP 在所有反应中结合到复合物上的量都是相同的，表明没有发生 GTP 水解（^{32}P 标记是在 γ-磷酸上，如果发生水解应该释放出磷酸）。最后，实验 3 表明 50S 核糖体的加入释放了 GTP，但不释放 fMet-tRNA$_f^{Met}$，后者要逗留到蛋白质合成的起始。在本章的后面将进一步讨论这一问题。

1973 年，John Fakunding 和 John Hershey 用标记的 IF2 和 fMet-tRNA$_f^{Met}$进行体外实验，结果表明它们都能结合到 30S 核糖体亚基上，且这种结合无需 GTP 水解。他们用^{3}H 标记 fMet-tRNA$_f^{Met}$，用[^{32}P] ATP 对 IF2 磷酸化，磷酸化的 IF2 保留了全部活性。然后将这些成分在 GTP 或不可水解的 GTP 类似物 **GDPCP** 存在下与 30S 核糖体亚基混合。GDPCP 的 β 磷酸和 γ 磷酸之间有一个亚甲基键（—CH_2—），而正常 GTP 在此处是一个氧原子，这是它不能水解成 GDP 和磷酸的原因。将所有这些成分混合后，Fakunding 和 Hershey 用蔗糖梯度超速离心分离各起始复合物。图 17.11 给出了实验结果。所有的标记 IF2 和大量的 fMet-tRNA$_f^{Met}$与 30S 核糖体亚基一起移动，表明起始复合物的形成。在 GTP 或 GDPCP 存在条件下有相同的结果，证明 GTP 水解不是 IF2 或 fMet-tRNA$_f^{Met}$对复合物结合所必需的。实际上，IF2 可以在 GTP 缺乏条件下结合到 30S 亚基上，但只发生在非天然的高浓度条件下。

这种实验也使 Fakunding 他们能够估计 30S 亚基、IF2 和 fMet-tRNA$_f^{Met}$两两间的结合量。加入越来越多的 IF2 产生一个饱和曲线，曲线在每个 30S 亚基结合 0.7mol IF2 时达到稳定。这个数字接近 1.0，有些 30S 亚基可能对结合 IF2 不敏感，可以认为真正的化学比是 1∶1。此外，在饱和 IF2 浓度，有 0.69mol fMet-tRNA$_f^{Met}$结合 30S 亚基，这几乎是 IF2 的结合量，所以 fMet-tRNA$_f^{Met}$结合的化学比看来也是 1∶1。我们将看到，IF2 最终会从起始复合物上释放出来再循环，结合另一个 fMet-tRNA$_f^{Met}$形成另一个复合物。以这种方式，它确实催化性地发挥作用。

接着，Fakunding 和 Hershey 研究了在缺乏 fMet-tRNA$_f^{Met}$及 AUG 时，IF1 和 IF3 对 IF2 结合到 30S 核糖体亚基上的影响。将标记的 IF2 与不同组合的因子及 30S 核糖体混合，然后离心混合物，看 IF2 是否已结合。没有 fMet-tRNA$_f^{Met}$和 AUG 时，IF2 对 30S 颗粒的结合不稳定，在其他因子和 GTP 缺乏时，IF2 因子自己不能结合到 30S 颗粒上（图 17.12）。同样，当 IF2 与 IF1 或 IF3 一起时，也不能发生结合。只有当 3 个因子都存在时，IF2 才能结合。这提示 3 个因子合作性地结合到 30S 亚基上。3 个因子都结合是形成 30S 起始复合物的第一

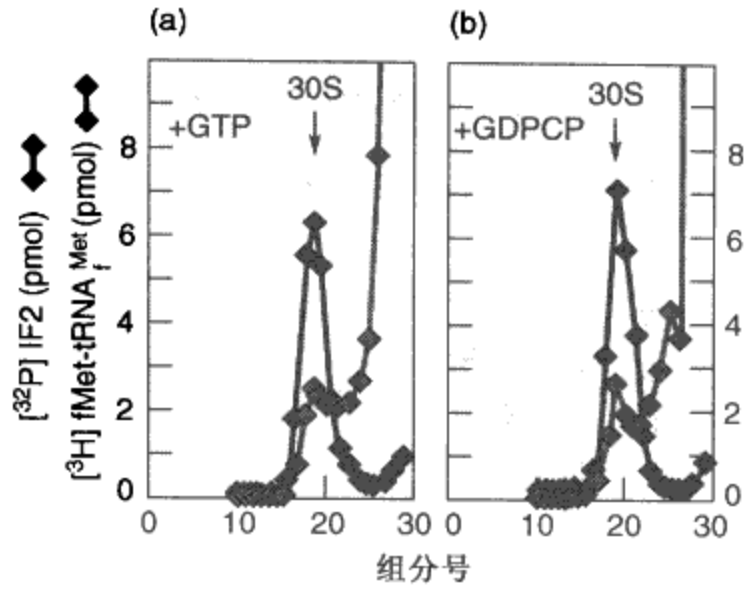


图 17.11　GTP 或 GDPCP 存在时 30S 起始复合物的形成。将 [^{32}P] IF2、[^{3}H] fMet-tRNA$_f^{Met}$、AUG、mRNA 替代物和 30S 核糖体亚基与（a）GTP 或（b）GTP 的非水解类似物 GDPCP 混合。用蔗糖梯度超速离心，分析每一梯度组分的放射性 IF2（蓝色）和 fMet-tRNA$_f^{Met}$（红色）。在 GTP 或者 GDPCP 存在情况下，IF2 和 fMet-tRNA$_f^{Met}$ 都能很好地结合 30S 核糖体亚基。（*Source*：Adapted from Fakunding, J. L. and J. W. B., Hershey, The interaction of radioactive initiation factor IF2 with ribosomes during initiation of protein synthesis. *Journal of Biological Chemistry* 248：4208，1973.）

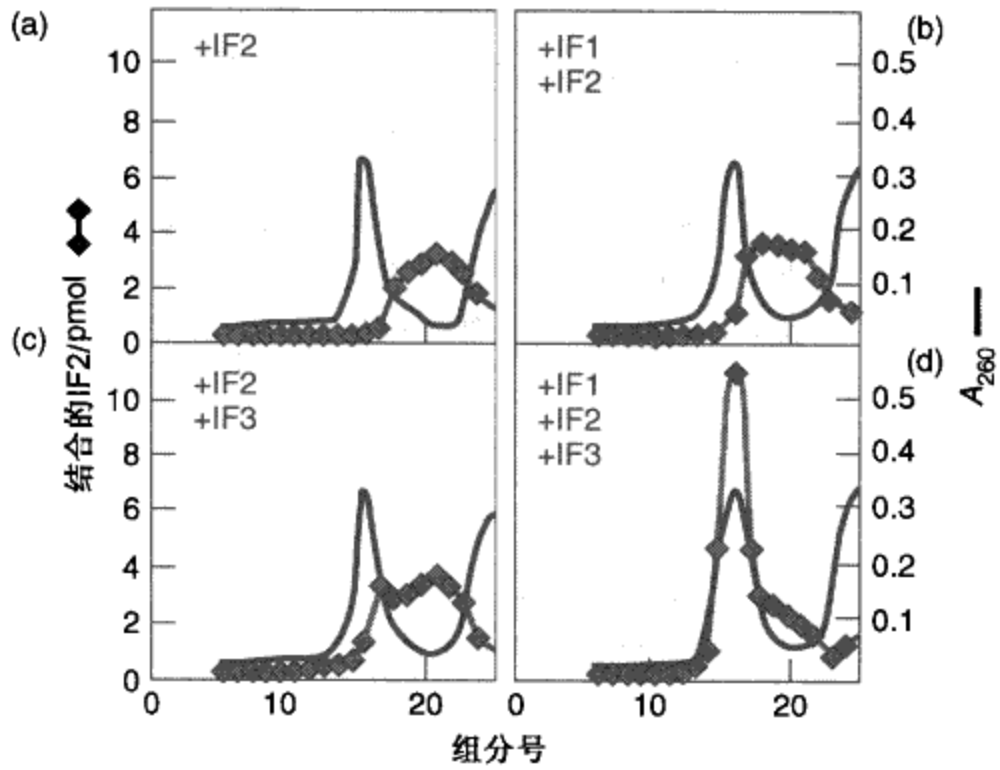


图 17.12　起始因子对 30S 核糖体亚基协同结合。将 30S 核糖体与 GTP、[^{32}P] IF2 和另外两种因子的不同组合混合，通过蔗糖梯度超速离心检测 IF2 的结合。因子组合为：（a）只有 IF2；（b）IF2 + IF1；（c）IF2 + IF3；（d）全部 3 个起始因子。蓝色，30S 核糖体，A_{260} 检测；红色，[^{32}P] IF2。在体外全部 3 个起始因子都存在时，IF2 才能结合。（*Source*：Adapted from Fakunding, J. L. and J. W. B. Hershey, The interaction of radioactive initiation factor IF2 with ribosomes during initiation of protein synthesis. *Journal of Biological Chemistry* 248：4208，1973.）

步。一旦结合，起始因子就能指导 mRNA 和 fMet-tRNA$_f^{Met}$ 的结合，产生完整的 30S 起始复合物，该复合物由 30S 核糖体亚基加上各一分子的 mRNA、fMet-tRNA$_f^{Met}$、GTP、IF1、IF2 和 IF3 组成。

小结　IF2 是促进 fMet-tRNA$_f^{Met}$ 与 30S 起始复合物结合的主要因子，另外两个起始因子起重要的辅助作用。在 IF2 的生理浓度下，GTP 是 IF2 结合所必需的，但在整个过程中 GTP 并不水解。完整的 30S 起始复合物包括一个 30S 核糖体亚基加上各一分子的 mRNA、fMet-tRNA$_f^{Met}$、GTP、IF1、IF2 和 IF3。

70S 起始复合物的形成

要发生延伸，50S 核糖体亚基必须结合 30S 起始复合物形成 **70S 起始复合物**（70S initiation complex）。在这个过程中，IF1 和 IF3 从复合物离解。当 IF2 离开复合物时，GTP 水解成 GDP 和磷酸盐。我们将看到 GTP 水解不能驱动 50S 核糖体亚基的结合，而是驱使 IF2 释放，否则 IF2 会干扰有活性 70S 起始复合物的形成。

已经知道 GTP 是 30S 起始复合物的一部分，在 50S 核糖体亚基结合复合物时 GTP 会被去除。怎样去除的呢？1972 年，Jerry Dubnoff 和 Umadas Maitra 证明 IF2 具有核糖体依赖性 GTPase 活性，使 GTP 水解成 GDP 和无机磷酸盐（Pi）。将 [γ-^{32}P] GTP 与盐洗过的核糖体（去除起始因子），或 IF2，或核糖体 + IF2 分别混合，对释放的 $^{32}P_i$ 作图。图 17.13 表明，核糖体或 IF2 单独都不能水解 GTP，但两者合起来则可以。因此，IF2 和核糖体一起组成了 GTPase。对 30S 起始复合物分析表明，30S 核糖体亚基不能以这种方式补充 IF2，因为直到 50S 颗粒加入到复合物后，GTP 才水解。

GTP 水解的作用是什么呢？Fakunding 和 Hershey 的 IF2 标记实验表明，GTP 水解是从核糖体上去除 IF2 所必需的。用标记的 IF2、fMet-tRNA$_f^{Met}$ 和 GDPCP 或 GTP 形成 30S 起始复合物，再加入 50S 亚基，然后高速离心混合物，观察哪个组分保留在 70S

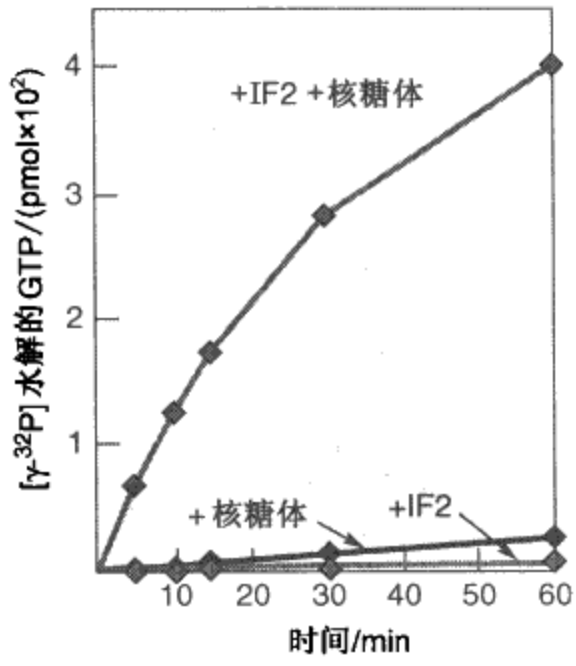


图 17.13　IF2 的核糖体依赖性 GTPase 活性。Dubnoff 和 Maitra 测量了在 IF2（绿色）存在时，标记的无机磷酸盐从［γ-^{32}P］GTP 的释放。核糖体（蓝色），IF2＋核糖体（红色）。核糖体和 IF2 共同水解 GTP。（*Source*：Adapted from Dubhoff, J. S., A. H. Lockwood, and U. Maitra, Studies on the role of guanosine triphosphate in polypeptide chain initiation in *Escherichia coli*. *Journal of Biological Chemistry* 247：2878，1972.）

起始复合物上。GDPCP 使 IF2 和 fMet-tRNA$_f^{Met}$都保留与 70S 复合物的结合，相反，GTP 使 IF2 游离，而使 fMet-tRNA$_f^{Met}$保留在 70S 复合物上（图 17.14），这证明 GTP 水解是 IF2 离开核糖体所必需的。

从图 17.14 中还可看出，在 GTP 存在时，结合到 70S 起始复合物上的 fMet-tRNA$_f^{Met}$比 GDPCP 存在时要多。这暗示了 IF2 的催化功能。GTP 水解是从 70S 起始复合物释放 IF2 所必需的，IF2 因而能将另一个 fMet-tRNA$_f^{Met}$结合到另一个 30S 起始复合物上，这种再循环构成了催化活性。但是，如果 GTP 未能水解，IF2 因子仍黏附在 70S 复合物上，不再循环，因而只能定量地发挥作用。Maitra 及同事在 IF2 浓度逐渐增加的情况下，测定［^{3}H］fMet-tRNA$_f^{Met}$对核糖体的结合得到了同样的结果。表 17.3 显示随着 IF2 浓度逐渐升高，越来越多的 fMet-tRNA$_f^{Met}$结合到 30S 核糖体上。然而，当他们加入 50S 核糖体亚基以及 30S 亚基时，也得到同样的效果。这可允许 GTP 水解释放 IF2，使其催化性地而不是化学定量性地执行功能。

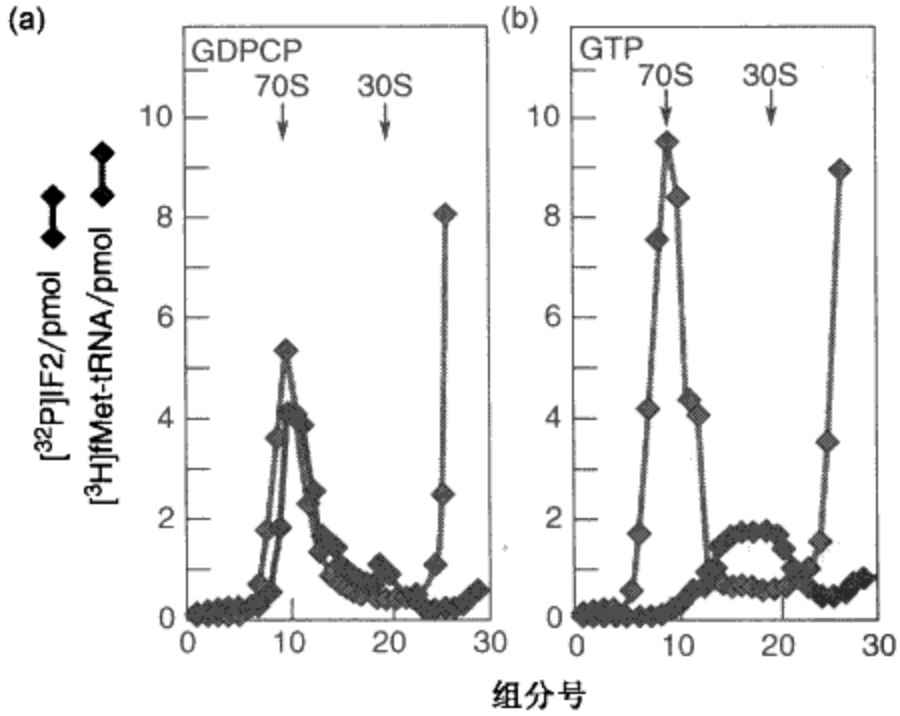


图 17.14　GTP 水解对 IF2 从核糖体释放的影响。Fakunding 和 Hershey 将［^{32}P］IF2（蓝色）、［^{3}H］fMet- tRNA$_f^{Met}$（红色）和 30S 核糖体亚基混合形成 30S 起始复合物。在 GDPCP（a）或 GTP（b）存在时加入 50S 核糖体亚基，用蔗糖梯度超速离心法（图 17.11）分析复合物。（*Source*：Adapted from Fakunding, J. L. and J. W. B. Hershey, The interaction of radioactive initiation factor IF2 with ribosomes during initiation of protein synthesis. *Journal of Biological Chemistry* 248：4210，1973.）

表 17.3　起始因子在 fMet-tRNA$_f^{Met}$ 对核糖体结合中的作用

	[^{3}H]fMet-tRNA$_f^{Met}$/pmol 结合到：	
加入的因子	30S	30S+50S
无	0.8	0.5
IF3(0.5pmol)	0.8	0.7
IF1(0.15μg)+ IF3(0.15μg)+ IF2(0.55pmol)	1.0	5.7
IF2(1.65pmol)	1.0	
IF2(8.8pmol)	2.2	
IF2(16.5pmol)	4.2	

Source: Dubnoff, J. S., A. H. Lockwood, and U. Haitra, Studies on the role of guanosine triphosphate in polypeptide chain initiation in Escherichia coli, Journal of Biological Chemistry 247: 2886, 1972. Copyright © 1972. The American Society for Biochemistry & Molecular Biology, Bethesda, MD. Reprinted with permission.

引发核糖体翻译也需要 GTP 水解吗？显然不是。Maitra 及同事通过凝胶过滤法从 30S 起始复合物上去除 GTP，发现这些复合物可以接受 50S 亚基并形成肽键，在这个过程中 GTP 没有水解，用 GDPCP 进行的相似过程有同样结果。因此，至少在这些实验条件下，GTP 水解不是产生有活性 70S 起始复合物的前体条件。GTP 水解的真正作用是将 IF2 和 GTP 本身从 70S 起始复合物上去除，使其行使连接氨基酸形成蛋白质的功能。

小结　50S 亚基结合 30S 复合物形成 70S 起始复合物后 GTP 被水解。这种 GTP 水解是由 IF2 连同 50S 核糖体亚基进行的，水解的目的是从复合物上释放 IF2 和 GTP，从而使多肽链的延伸能够开始。

细菌中的翻译起始总结

图 17.15 总结了已学习过的细菌的翻译起始，包括下列特征。

1. 在 IF1 的影响下，70S 核糖体解离为 50S 和 30S 亚基。

2. IF3 对 30S 亚基的结合抑制了核糖体亚基的再结合。

3. IF1、IF2 和 GTP 结合在 IF3 旁边。

4. mRNA 和 fMet-tRNA$_f^{Met}$ 结合形成 30S 起始复合物。这两个组分可以明显地相互结合，但 IF2 促进 fMet-tRNA$_f^{Met}$ 结合，IF3 促进 mRNA 结合，在每种情形下，其他起始因子也都提供协助。

5. 50S 亚基的结合使 IF1 和 IF3 丢失。

6. IF2 从复合物上解离，同时 GTP 水解，产物是 70S 起始复合物，可以开始延伸了。

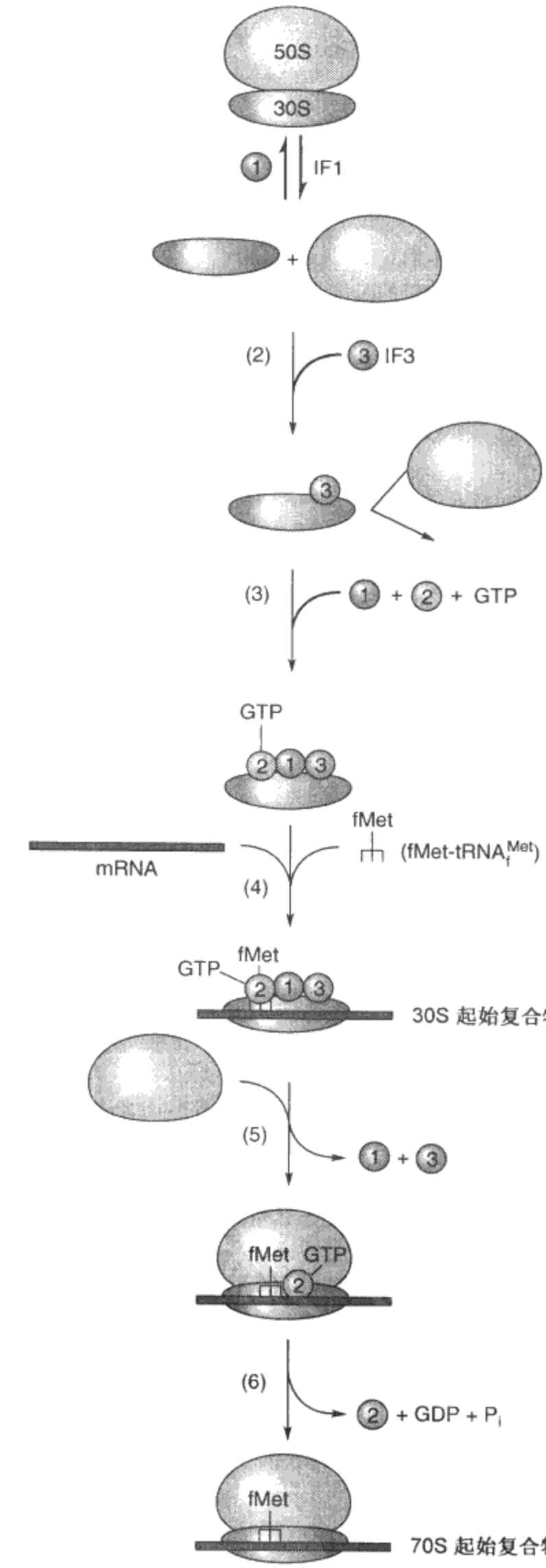


图 17.15　细菌翻译起始总结。步骤 1～6 见正文，步骤 2 和 3 可能结合在体内。

17.2 真核生物翻译的起始

有几个特征使真核生物的翻译起始不同于细菌的翻译起始。第一，真核生物的起始从甲硫氨酸，而不是 *N*-甲酰甲硫氨酸开始。但起始 tRNA 与将甲硫氨酸加到多肽链内部的 tRNA（$tRNA_m^{Met}$）不同。起始 tRNA 携带的是非甲酰化甲硫氨酸，不适合叫做 $tRNA_f^{Met}$，而常被称作 **$tRNA_i^{Met}$** 或者 **$tRNA_i$**。第二个主要区别是，真核生物的 mRNA 没有 Shine-Dalgarno 序列来指示核糖体从哪儿开始翻译，大多数真核生物 mRNA 的 5′端具有帽子结构（第 15 章），以此指示起始因子结合并开始搜索起始密码子。这种不同的起始机制及其所要求的起始因子将是本节讨论的主题。

起始的扫描模型

多数细菌 mRNA 是多顺反子，含有多个基因或顺反子的遗传信息。每个顺反子都有自己独立的起始密码子和核糖体结合位点。但除了某些病毒的转录物外，多顺反子 mRNA 在真核生物中很少见。因此，真核细胞经常面临在转录的 5′端寻找起始密码子的任务，它们识别 5′端帽子结构，然后沿 5′→3′方向**扫描**（scanning）mRNA，直到遇见起始密码子，如图 17.16 所示。

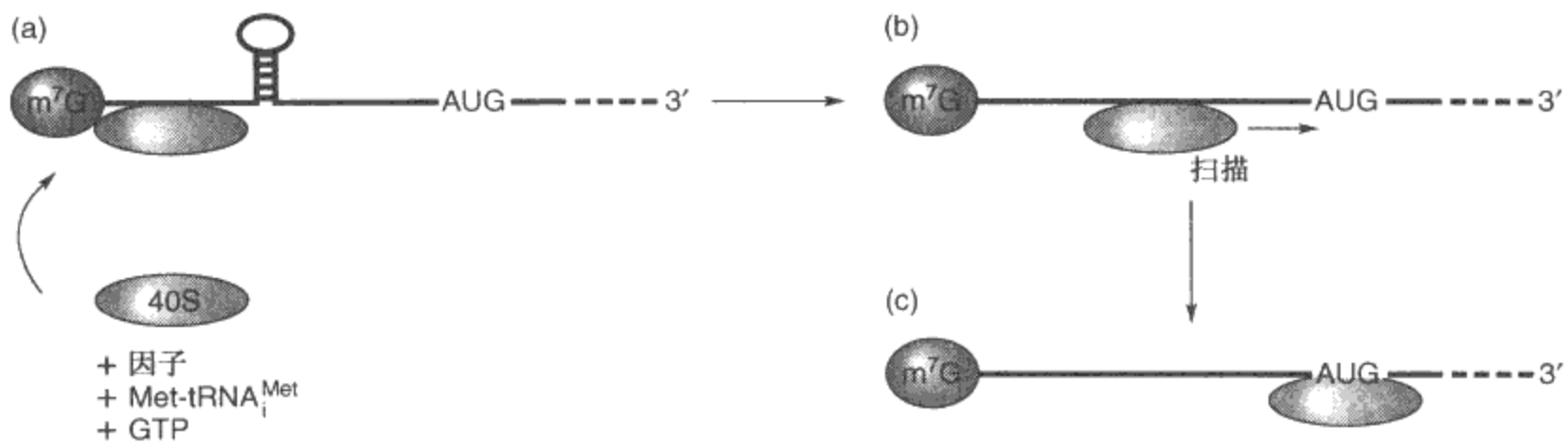


图 17.16 简单的翻译起始“扫描模型”。(a) 40S 核糖体亚基与起始因子、$Met\text{-}tRNA_i^{Met}$ 和 GTP 一起识别 mRNA 5′端的 m^7G 帽子（红色），使核糖体亚基在 mRNA 该末端结合。为简单起见，其他组分被省略了。(b) 40S 亚基沿着 mRNA 向 3′端扫描，搜索起始密码子。在此过程中解开一个茎环结构。(c) 核糖体亚基定位于 AUG 起始密码子处并停止扫描。现在 60S 核糖体亚基可以加入复合物，起始可以发生了。

基于以下 4 点考虑，Marilyn Kozak 于 1978 年首先建立了“扫描模型”：①没有证据表明真核生物的翻译起始像多顺反子 mRNA 一样从内部 AUG 开始；②起始不在离 mRNA 5′端固定的距离发生；③在研究过的所有前 22 个真核生物 mRNA 中，帽子结构下游的第一个 AUG 都被用作起始；④mRNA 5′端的帽子结构促进起始。在本章后面我们将看到扫描模型的更有力的证据。

最简单版本的扫描模型是核糖体识别它遇到的第一个 AUG 并在那里起始翻译。但是，对 699 个真核生物 mRNA 的调查显示，5%～10%的 mRNA 第一个 AUG 不是起始位点，其核糖体在遇到正确的 AUG 并起始翻译之前跳过了一个或更多的 AUG。由此产生了一个问题：什么东西规定正确 AUG 区别于错误 AUG? Kozak 研究了起始 AUG 周围的序列，发现共有序列是 CCRCCAUGG，R 是嘌呤（A 或 G），加下画线的是起始密码子。

如果这真是最佳序列，那么就会降低突变效率。为验证这一假说，Kozak 在克隆大鼠前胰岛素原基因中围绕起始密码周围的碱基系统地进行了突变。用人工合成的含有 ATG 的寡核苷酸取代正常的起始 ATG，对这个起始区域引入突变，然后将突变基因置于 SV40 病毒启动子的控制之下，转入猴细胞（COS）中，然后用 ^{35}S 标记的甲硫氨酸标记新合成的蛋白质，免疫沉淀前胰岛素原，电泳、荧光检测（类似放射自显影技术）。最后，用密度计扫描荧光图，对胰岛素原进行定量。翻译起始得越好，合成的胰岛素原越多。在讨论过程中，我们称起始密码子为 AUG，即使突变是在 DNA 水平上发生的。

图 17.17 显示了部分结果，包括在－3 和＋4 位碱基的改变（AUG 中的 A 是＋1 位）。最佳起始发生在－3 位为 G 或 A，＋4 位为 G 时。相似的实验显示所有最佳起始都发生在 ACCAUGG 序列，这些要求通常称作 **Kozak 规则**（Kozak's rule）。

如果这是翻译起始的最佳序列，那么将其引到可读框之外并在正常起始密码子的上游，将会对核糖体扫描产生障碍，迫使它们在可读框外起始。这种情况发生得越多，产生的胰岛素原越少。Kozak

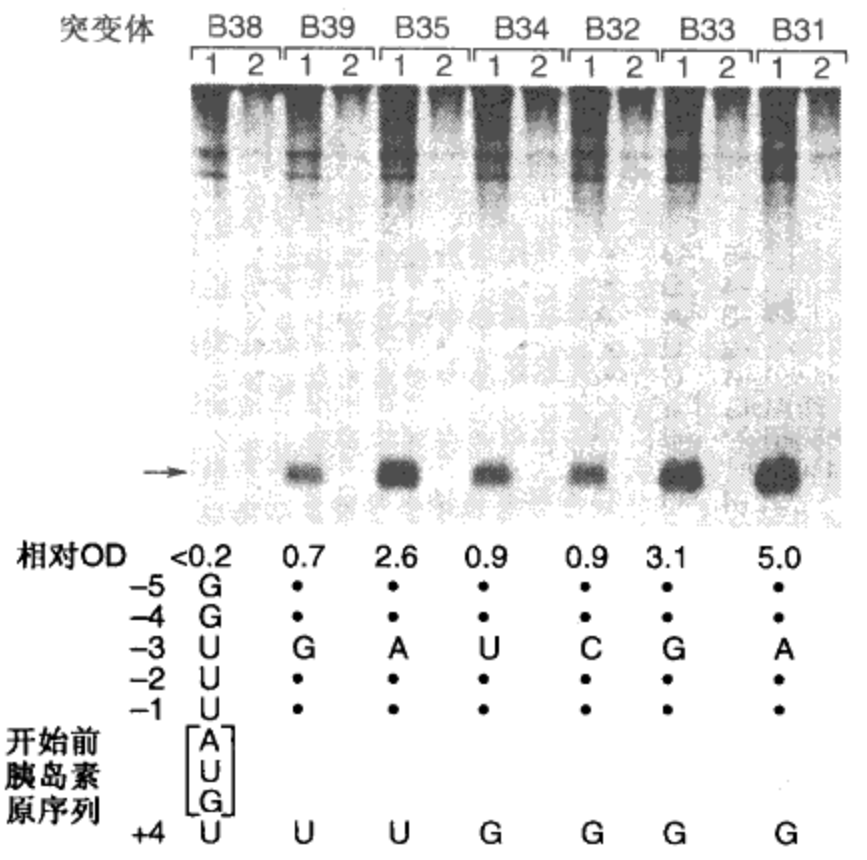


图 17.17　在起始 AUG 周围－3 和＋4 位单个碱基改变的影响。从 SV40 病毒启动子控制下的大鼠前胰岛素原克隆基因开始，Kozak 用人工合成的含 ATG 的寡核苷酸代替天然起始密码子，该 ATG 在 mRNA 中被转录为 AUG。然后她突变了－3 和＋4 位的核苷酸，如底部所示。将突变的基因转入 COS 细胞，在含有［^{35}S］标记的 Met 培养基中培养细胞，以便标记产生的胰岛素原。免疫沉淀法纯化胰岛素原，电泳、荧光摄影检测标记的蛋白质。此技术与放射自显影类似，电泳凝胶中充满荧光混合物，用来增强同位素（如^{35}S）散发的较弱放射性。左边箭头指示胰岛素原产物，荧光摄像测定胰岛素原条带的亮度，以每个条带下列出的相对 OD 值或光密度表示。最佳起始发生在－3 是嘌呤，＋4 是 G。胰岛素原是前胰岛素原基因的产物，在前胰岛素原翻译期间，N 端的“信号肽”被切除，产生胰岛素原。信号肽指导生长的多肽连同核糖体、mRNA 一起到内质网（ER），以确保多肽进入 ER 并分泌到细胞外。所有序列以 mRNA 形式给出。（*Source*：Kozak，M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* 44（31 Jan 1986）p. 286，f. 2. Reprinted by permission of Elsevier Science.）

用两个 AUG 的 A 相距 8nt 的序列（AUGNCACC-AUGG）进行了实验。注意，下游的 AUG 是最佳位置。因此，如果核糖体未在上游先起始而到达这里，那么起始是很容易的。如图 17.18 所示，突变体 F10 无上游 AUG，从正常 AUG 的起始信号很强，如所预计的。突变体 F9 的上游有很弱形式的 AUG，其－3 和＋4 位都是 U，同样，对下游 AUG 的起始没有太多干扰。但是所有其他突变表现出对正常起始的强烈干扰，且干扰强度与上游 AUG 的前后序列有关。与最佳序列越相似，对下游 AUG 的起始干扰就越多。这正是扫描模型所预计的。

上游 AUG 具有良好的序列环境，但仍设法从下游 AUG 启动的天然 mRNA 会怎样呢？Kozak 注意到这些 mRNA 在两个 AUG 之间有框内终止密码子。她认为，下游 AUG 的起始实际上体现了核糖体的再起始，核糖体已经在上游起始密码子处起始，在终止密码子处停止，然后继续扫描另一个起始密码子。为了阐明两个 AUG 之间的终止密码子的效果，Kozak 构建了另一套带有此终止密码子的系统，采用同样的方法检测。在这种情形下，只要下游 ATG 处于良好的序列环境，大量起始就可发生在下游 AUG 处。

注意，在同一可读框内的起始密码子和下游终止密码子决定了**可读框**（open reading frame，ORF）的边界。这种 ORF 可能编码一个蛋白质，它在体内是否被翻译是另外一回事。更多的实验揭示了下游 ORF 有效再起始的另一个要求：上游 ORF 必须是短的。在已研究的带有全长上游 ORF 的双顺反子的每个例子中，下游 ORF 的再起始效率都非常低。可能核糖体在完成一个长 ORF 的翻译后，再起始所需的因子已散去，所以它忽略了第二个 ORF。

为了严密地论证上游 AUG 比下游 AUG 优先的假说，Kozak 构建了带有多个重复的大鼠前胰岛素原顺反子起始区序列的 mRNA，通过分离所得蛋白质来检测实际的翻译起始位点。电泳测定其大小，由此得知核糖体利用哪个起始位点来合成这些蛋白质。图 17.19 的结果表明，在每种情况下最远的上游 AUG 都被使用了，再次与扫描模型一致。

mRNA 二级结构对起始效率有何影响？mRNA 的发夹结构对起始效率既有正面影响也有负面影响。Kozak 证明弱序列环境的 AUG 下游 12～15nt 的茎环（stem loop）结构能阻止 40S 核糖体亚基跳过该起始位点。发夹结构可能使核糖体亚基在 AUG 停留足够长的时间使起始发生。二级结构也有负效应，Kozak 检测了 mRNA 前导序列的两个不同茎环结构的影响（图 17.20a）。一个相对较短，具有－30kcal/mol的双链形成自由能（或稳定性）；另一个长得多，具有－62kcal/mol 的自由能。将这些茎环引入氯霉素乙酰转移酶（CAT）基因前导序列的不同位点，然后体外转录，并在^{35}S 标记的甲硫氨酸存在时翻译其转录物。最后，电泳检测 CAT 蛋白、荧光成像检测。图 17.20b 的结果表明：帽子下游 52nt

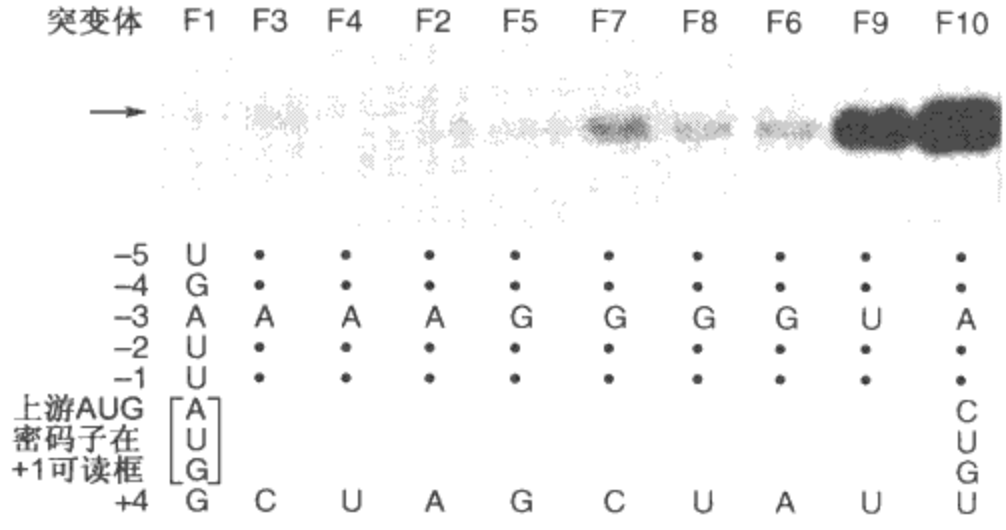


图 17.18　上游“障碍”AUG 序列环境的影响。Kozak 构建了一个含有大鼠前胰岛素原转录物正常 AUG 起始密码子的质粒，转录物之前是可读框外的 AUG。然后在上游 AUG 周围的−3 和+4 位点进行突变，分析对胰岛素原合成的影响，左边箭头指示正确起始的胰岛素原。上游 AUG 前后的序列环境越好，作为障碍物纠正下游起始就越好。提供的所有序列与在 mRNA 中的相同。（*Source*：Kozak，M.，Point mutations define a sequence flanking the AUG initiation codon that modulates translation by eukaryotic ribosomes. *Cell* 44（31 Jan 1986）p. 288，f. 6. Reprinted by permission of Elsevier Science.）

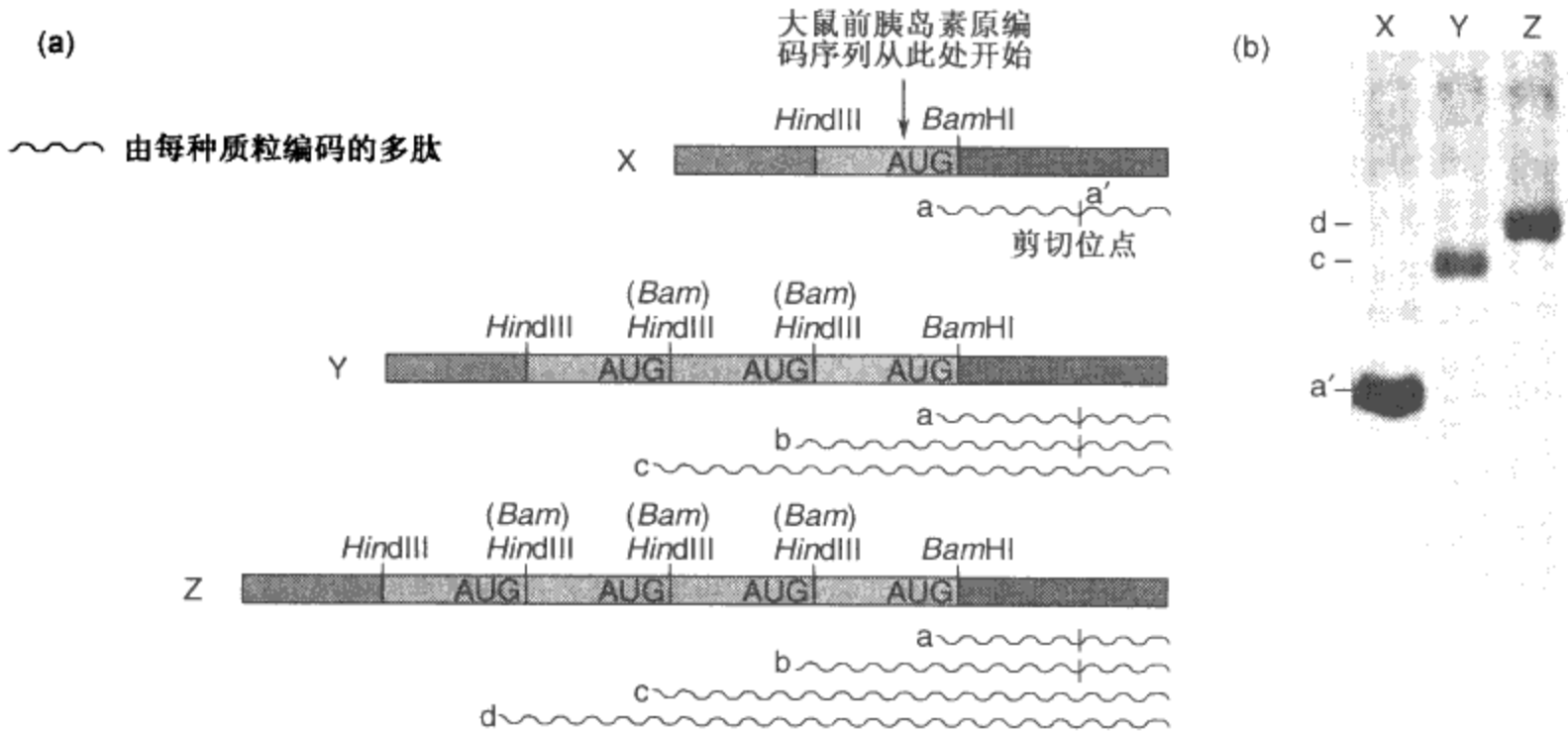


图 17.19　多重翻译起始位点的影响。（a）构建。Kozak 构建了 3 个结构：X、Y 和 Z，带有 1、3 和 4 个拷贝的大鼠前胰岛素原起始位点。此位点包含在 *Hin*dIII-*Bam*HI DNA 片段内，能够首尾相连，如 Y 和 Z 所示，起始位点的重复拷贝前后相连。每一结构下面的波浪线表示编码的可能的蛋白质。（b）结果。Kozak 将这些结构转入哺乳动物细胞，使之转录、翻译。然后，免疫沉淀法检测含胰岛素原的多肽，电泳、荧光显色检测。真实的胰岛素原（a′）是结构 X 的产物，它只有正常的起始位点，信号肽能被切除。然而，结构 Y 有 3 个起始密码，产物只有 c，是从最远的上游密码子起始的。同样结构 Z 有 4 个起始密码，产物只有 d，是从最远的上游密码起始的。c 或 d 的信号肽不能被切除，因为它在蛋白质的内部。（*Source*：Kozak，M.，Translation of insulin-related polypeptides from messenger RNAs with tandemly reiterated copies of the ribosome binding site. *Cell* 34（Oct 1983）p. 975，f. 4. Reprinted by permission of Elsevier Science.）

的−30kcal 茎环结构不能干扰翻译，即使它包括起始 AUG，而帽子下游 12nt 的−30kcal 茎环结构强烈地抑制翻译，可能是因为它干扰了 40S 核糖体亚基和各因子在帽子处的结合。此外，在帽子下游 71nt 处的−62kcal 茎环能彻底阻止 CAT 蛋白的产生。

为什么具有稳定发夹的结构不能被翻译呢？最简单的解释是非常稳定的茎环阻止了正在扫描的 40S 核糖体亚基的通过，使之不能通过起始密码子。这种影响只在顺式结构（同一个分子上）中被观察到。当结构 3 和 4（或 3 和 1）合在一起检测时，结构 3（泳道 4 和 6）产生的线性 mRNA 上可发生翻译。这表明不可翻译结构在某种形式上对翻译系统是无害的。

结构 2 可被很好地翻译，即使其起始密码子被隐藏在发夹结构内，这一事实提示核糖体亚基和起始因

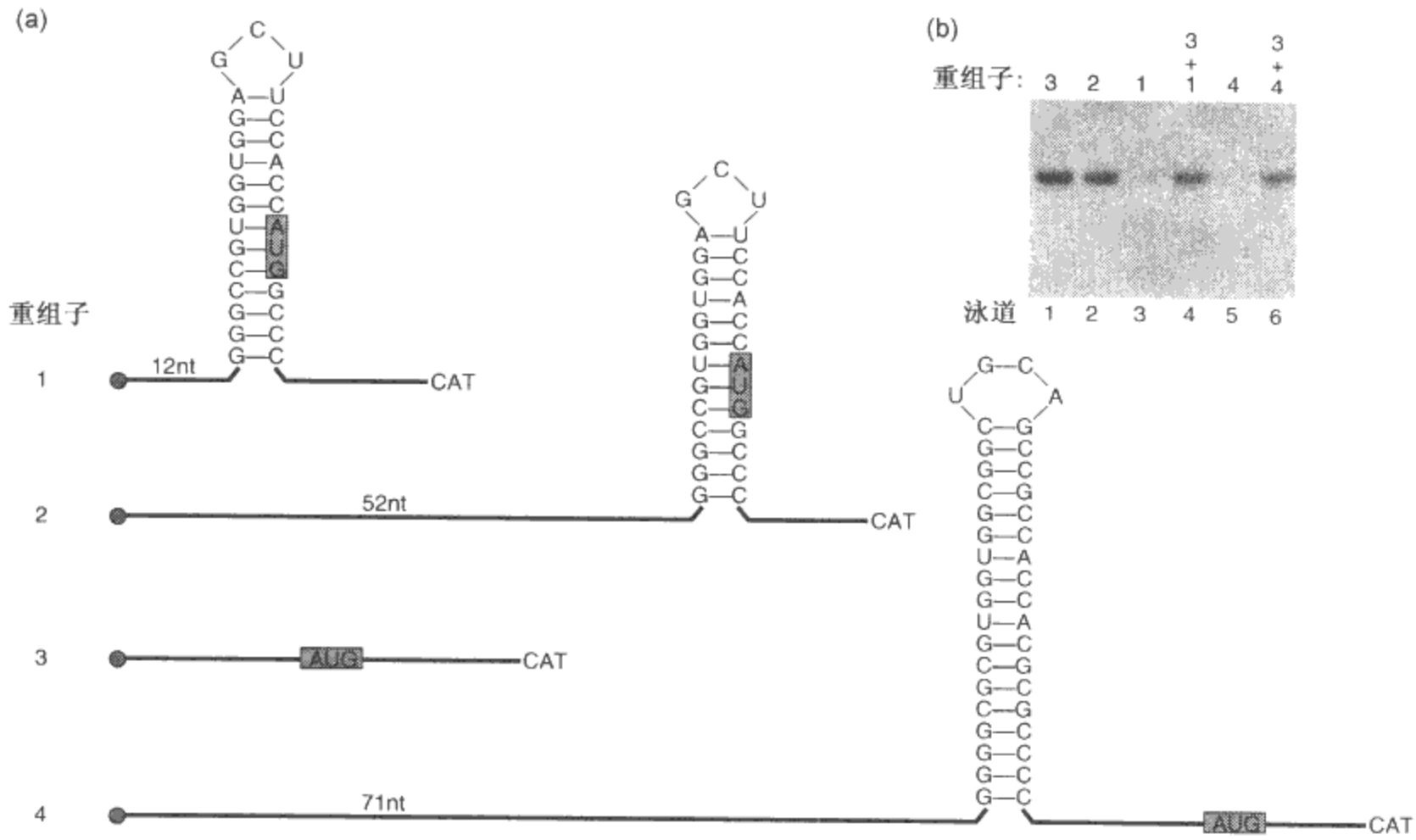


图 17.20　mRNA 前导序列的二级结构对翻译效率的影响。(a) mRNA 结构。Kozak 合成了图中的前导序列结构，红色是帽子结构，起始密码子由绿色突出，每一结构的 3′端连着 CAT 的 ORF。(b) 体外翻译结果。Kozak 在体外转录了酶结构，在兔网状细胞提取物中加有[^{35}S]甲硫氨酸来翻译 mRNA。电泳标记的蛋白质，荧光成像检测。靠近帽子（结构 1）的短发夹干扰了翻译间的长发夹也同样干扰了翻译。(*Source*：Kozak, M.，Circumstances and mechanisms of inhibition of translation by secondary structure in eukaryotic mRNAs. *Molecular and Cellular Biology* 9 (1989) p. 5136, f. 3. American Society for Microbiology.)

子能够打开一定量的双链 RNA，这和 Kozak 最初的扫描模型所预计的一样（图 17.16）。然而，正如我们刚才所见，这种解链能力有局限性，结构 4 中的长发夹有效地阻止了核糖体亚基到达起始密码子。

40S 核糖体亚基是怎样识别 AUG 起始密码子的呢？Thomas Donahue 及同事阐明了起始 tRNA ($tRNA_i^{Met}$) 起关键作用。他们把 4 个酵母 $tRNA_i^{Met}$ 中的一个反密码子变成 3′-UCC-5′，使它识别 AGG 而不是 AUG。然后，将带有多种突变起始密码子的 *his4* 基因转入 *his4*$^-$ 酵母菌株。图 17.21a 显示在起始密码子处携带 AGG 密码子的 *his4* 基因能够维持酵母生长，而其他替代的起始密码子都不行，可能是因为它们不能与已发生改变的起始 tRNA 的 UCC 反密码子配对。在另一个实验中，研究者把第二个 AGG 放在起始位点的 AGG 上游 28nt 处，以其改变可读框，发现这种结构不能维持生长。此结果符合扫描模型，如图 17.21b 所示。起始 tRNA 以及本例中的 UCC 反密码子结合到 40S 核糖体亚基上，该复合物扫描 mRNA，搜索第一个起始密码子（在这里是 AGG）。由于第一个 AGG 在 *his4* 编码区的可读框外，翻译将在错误的可读框发生，很快遇到终止密码子而提前终止。

扫描模型有一些明显的例外。其中最清楚的是细小核糖核酸病毒（picornavirus）的多顺反子 mRNA，如脊髓灰质炎病毒，它们缺乏帽子结构。在这些例子中，核糖体不必扫描可以很容易地在内部起始密码子处进入。少数有帽子结构的细胞 mRNA也有**内部核糖体进入序列**（internal ribosome entry sequence，IRES），可以直接吸引核糖体而不用帽子帮助，本章后面将更详细地讨论这一现象。

小结　通常真核 40S 核糖体亚基与起始 tRNA ($tRNA_i^{Met}$) 一起，通过结合到 mRNA 的 5′帽子并扫描下游直到在合适的序列环境中找到第一个 AUG，定位正确的起始密码子。最好的序列环境是 AUG 中的 A 在+1、嘌呤在−3 位、G 在+4 位。有5%～10%的情况，核糖体亚基将跨过第一个 AUG，继续搜寻更合适的 AUG。有时核糖体明显地在上游的 AUG 处起始，翻译一个短的 ORF，然后继续扫描并在下游的一个 AUG 处再起始。这种机制只在短的上游 ORF 起作用。靠近 mRNA 5′端的二级结构对起始有正的或负的

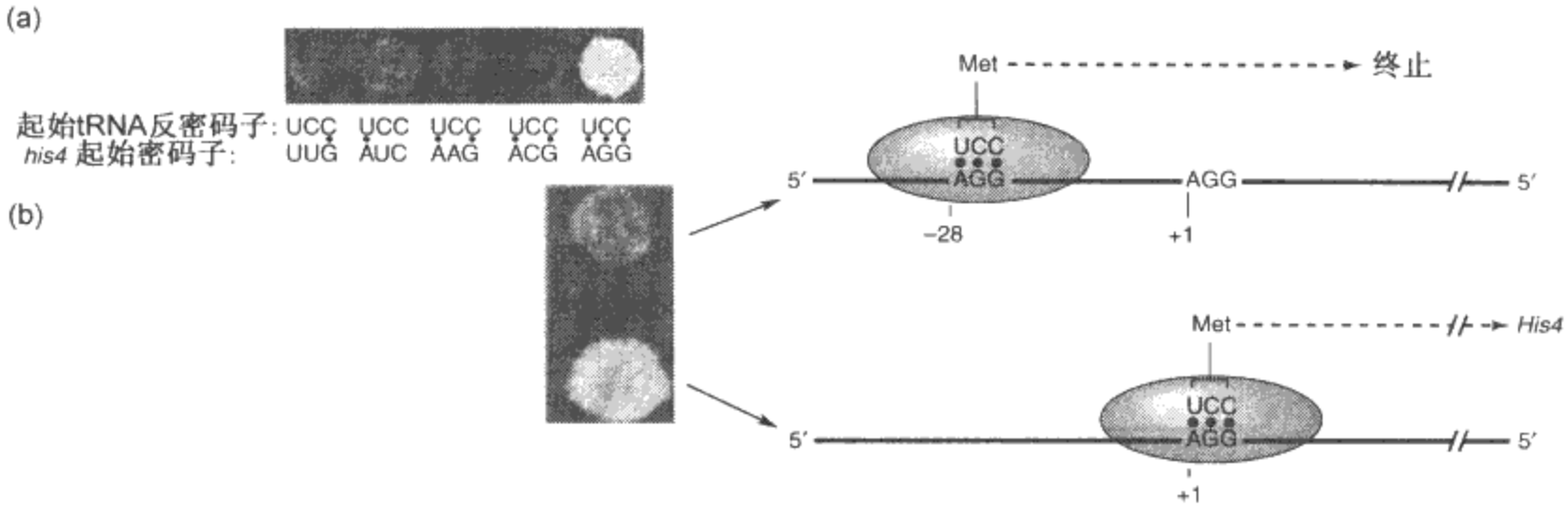


图 17.21　起始 tRNA 在扫描中的作用。（a）带有改变了反密码子的起始 tRNA 能够识别互补的起始密码子。Donahue 及同事把酵母起始 tRNA 中的一个反密码子突变为 3′-UCC-5′，然后，用高拷贝的酵母载体将编码该突变 tRNA 的基因转入 *his4*⁻酵母菌株。然后改变 *his4* 基因的起始密码子为底部所列 5 种形式中的任一种，检测在组氨酸缺乏时突变酵母细胞的生长。当起始密码子是 AGG 时，能与起始 tRNA 的反密码子 UCC 配对，突变的 mRNA 能够被翻译，生长发生。（b）上游额外的 AGG 和可读框移位的影响。Donabue 和他的同事构建了一个 *his4*，在好序列环境下其−28 位（上）加了一个额外的 AGG，将其转入带有 UCC 反密码子的起始 tRNA 细胞中，检测细胞在缺乏组氨酸时的生长能力，发现其生长比没有上游 AGG（下）的细胞明显减慢。显然，扫描中的 40S 核糖体亚基与突变的 $tRNA_i^{Met}$ 一起，遇到第一个 AGG 并在那儿起始翻译，产生变短的 *his4* 产物。（*Source*：（a）Cigan，A. M.，L. Feng，and T. F. Donohne，$tRNA_i^{Met}$ functions in directing the scanning ribosomes to the start site of translation. *Science* 242（7 Oct 1988）p. 94，f. 1B & C（left）. Copyright © AAAS.）

影响。紧接 AUG 之后的发夹结构能迫使核糖体亚基暂停于 AUG，因而激发起始。在帽子结构与起始位点之间非常稳定的茎环结构能阻止核糖体亚基扫描，因而阻止起始。一些病毒和细胞 mRNA 含有 IRES，能吸引核糖体直接到 RNA 内部。

真核生物的起始因子

我们已经了解到细菌的翻译起始需要起始因子参与，真核生物也同样需要起始因子。可是如你所预期，真核系统比细菌系统要复杂。我们已经了解的其中一个额外复杂层次是扫描过程，需要一些因子识别 mRNA 5′端的帽子并与附近的 40S 核糖体亚基结合。在本节，我们将探讨真核生物起始不同阶段所参与的多种因子，我们还将了解到其中一些步骤是翻译过程调节的天然位点。

真核生物翻译起始概况

图 17.22 描述了真核生物翻译起始的概况，显示了参与起始的主要起始因子。注意，真核生物起始因子的第一个字母都是 e，代表“eukaryotic”。**eIF2** 与原核生物的 IF2 功能相似，负责把氨酰-tRNA（Met-$tRNA_i^{Met}$）结合到核糖体上。

eIF2 与 IF2 类似的另一方面是需要 GTP 完成这一工作。当 eIF2 从核糖体解离时，GTP 水解成 GDP，GTP 必须取代 eIF2 上的 GDP，使 eIF2 恢复功能。这一过程需要交换因子 **eIF2B**，使 eIF2 上的 GDP 与 GTP 交换。该因子也称作 **GEF**，表示鸟嘌呤核苷酸交换因子（guanine nucleotide exchange factor）。注意，在某个步骤发挥功能的所有因子都以相同数字标记。例如，起始氨酰-tRNA 的结合至少需要两个因子（eIF2 和 eIF2B）参与，两个因子都用数字 2 标注。尽管 IF2 和 eIF2 之间具有结构上的相似性，但两者并不同源，而是 IF2 与 eIF5B 同源，后面将进行探讨。

另一个与原核因子功能相似的真核因子是 **eIF3**，它结合 40S（小）核糖体亚基，并阻止其与 60S（大）亚基的重新结合，这类似于 IF3 的功能。**eIF4F** 是一个复杂的帽子结合蛋白，使 40S 核糖体颗粒结合到 mRNA 的 5′端。一旦 40S 颗粒结合到帽子上，就要求 **eIF1**（和 **eIF1A**）去扫描起始密码子。**eIF5** 没有细菌对应物，它促进 60S 核糖体亚基与 40S 起始复合物的结合，实际上 40S 起始复合物叫做 48S 复合物，因为除了 40S 核糖体亚基外，它还包含 mRNA 和许多因子，从而增加了沉降系数。像 eIF3 一样，**eIF6** 是另一个抗结合因子，它与 60S 核糖体亚基结合防止 40S 亚基的过早结合。

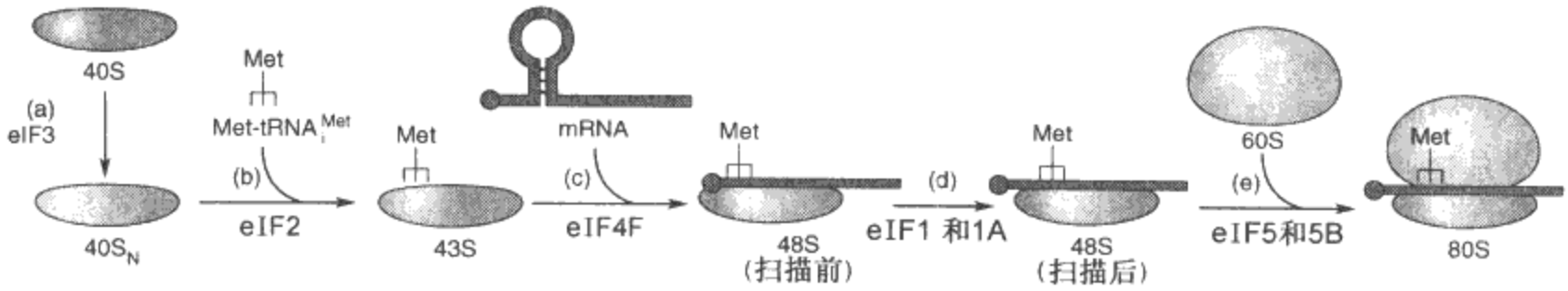


图 17.22　真核生物翻译起始小结。（a）eIF3 使 40S 核糖体亚基转变为 40S$_N$，抵抗与 60S 核糖体颗粒的结合，并准备接受起始氨酰-tRNA；（b）在 eIF2 的帮助下，Met-tRNA$_i^{Met}$ 结合 40S$_N$ 颗粒，形成 43S 复合物；（c）在 eIF4F 协助下，mRNA 结合到 43S 复合物，形成 48S 复合物；（d）eIF1 和 1A 因子促进对起始密码子的扫描；（e）eIF5 因子促进 eIF2-结合的 GTP 的水解，这是核糖体亚基加入的前提，eIF5B 有核糖体依赖性 GTPase 活性，能帮助 60S 核糖体颗粒结合 48S 复合物，形成 80S 复合物，准备好翻译 mRNA。

小结　真核生物起始因子有下列一般功能：eIF2 参与 Met-tRNA$_i^{Met}$ 与核糖体的结合。eIF2B 通过 GTP 取代 GDP 而激活 eIF2。eIF1 和 eIF1A 帮助扫描起始密码子。eIF3 结合 40S 核糖体亚基并阻止 40S 与 60S 亚基的重新结合。eIF4F 是个帽子结合蛋白，能使 40S 核糖体亚基结合到 mRNA 5′端。eIF5 促进 60S 核糖体亚基与 48S 复合物（40S 亚基＋ mRNA＋ Met-tRNA$_i^{Met}$）的结合。eIF6 结合 60S 亚基，阻止 60S 亚基与 40S 亚基的重新结合。

eIF4F 的功能　我们来看看真核生物翻译起始的主要特点：帽子结构的作用。我们已了解到帽子结构极大地促进了 mRNA 的翻译效率。这意味着某个因子能够识别 mRNA 5′端的帽子结构，帮助 mRNA翻译。1978 年，Nahum Sonenberg 及同事通过将蛋白质交联到修饰过的帽子上鉴定出一个帽子结合蛋白。做法是先将^{3}H-呼肠孤病毒 mRNA 的加帽核苷酸的核糖氧化，使其 2′-OH 和 3′-OH 变成活性二醛，然后将改变的 mRNA 与起始因子一起孵育。结合到被修饰帽子上的任何因子的自由氨基将共价地结合到其中一个活性乙醛上，这种结合可经还原作用而持久。交联之后，用 RNase 消化除了帽子之外的所有 RNA。对产物电泳，检测任何交联到标记帽子上的蛋白质大小。图 17.23 显示即使在低温条件下，有一条约 24kDa 的多肽可结合。在较高温度下，另外一对较高分子质量（50～55kDa）的多肽也能结合。然而，这些高分子质量多肽的结合不能被未标记的 m^{7}GDP 竞争，而对 24kDa 多肽的结合可以被帽子类似物竞争，提示 24kDa 多肽特异性地与帽子结合，而 50～55kDa 多肽是非特异性地结合。另一方面，GDP 与 50～55kDa 多肽竞争对 mRNA 的结合，但不与 24kDa 多肽竞争，提示这种大多肽可能是 GDP-结合蛋白，而不是帽-结合蛋白。

Sonenberg 及同事在他们对帽子结合蛋白研究的发现后，接着通过 m^{7}GDP-琼脂糖柱亲和层析纯化

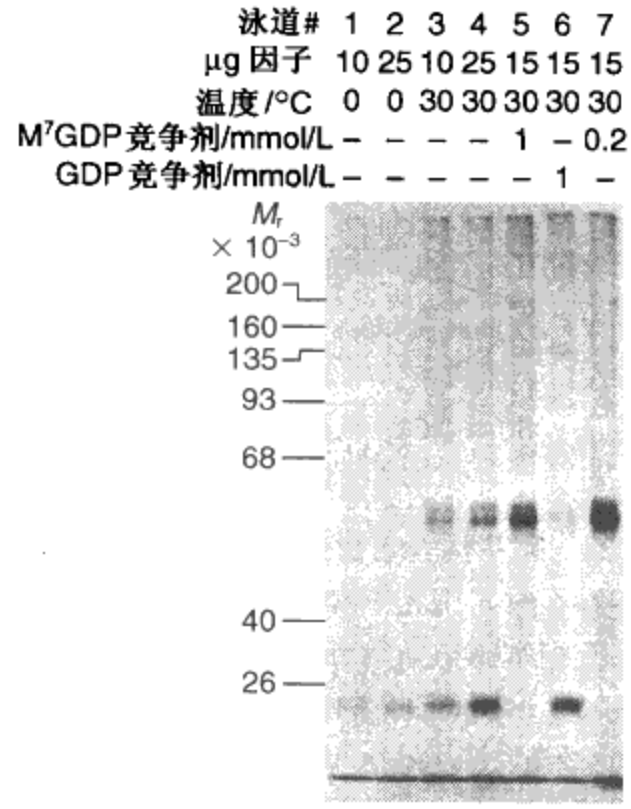


图 17.23　通过化学交联法鉴定帽子结合蛋白。将^{3}H-呼肠孤病毒 mRNA 加帽核苷酸的核糖，氧化成一个活性二醛。然后将起始因子和 mRNA 交联耦合任一帽子结合蛋白，这种耦合是通过帽子上的乙醛和蛋白质上一个自由氨基间的一个 Schiff 碱基来完成的。用 $NaBH_3CN$ 处理使共价键因变形而更持久。接着用 RNase 消化复合物，去除帽子以外的所有成分。电泳分离标记的帽子结合蛋白复合物，检测任何结合到帽子结构的多肽的大小。每一泳道的条件在图的顶部给出。发现 24kDa 条带的结合被 m^{7}GDP 竞争，但 50～55kDa 条带不能。（*Source*：Sonenberg，N.，M. A.，Morgan，W. C. Merrick，and A. J. Shatkin，A polypeptide in eukaryotic initiation factors that crosslinks specifically to the 5′-terminal cap in mRNA. *Proceedings of the National Academy of Science USA* 75（1978）p. 4844，f. 1.）

该蛋白质，然后将其加入无 HeLa 细胞的提取物中，证明它能促进有帽 mRNA 翻译，但不能促进无帽 mRNA 的翻译（图 17.24）。在两个实验中，他们用的都是病毒 mRNA，辛德毕斯（Sindbis）病毒 mRNA是加帽的，脑心肌炎病毒（encephalomyocarditis）mRNA 是无帽的（脑心肌炎病毒是一个类似于脊髓灰质炎病毒的细小核糖核酸病毒）。我们知道，细小核糖核酸病毒 mRNA 没有帽子结构，但它们有相应的机制确保其 mRNA 翻译。事实上它们正是利用其 mRNA 的无帽特性消除了宿主有帽 mRNA 的竞争，通过失活宿主的帽子结合蛋白，阻断宿主有帽 mRNA 的翻译。分子生物学家据此以脊髓灰质炎病毒感染细胞的提取物作为帽子结合蛋白分析系统，任何能恢复该细胞提取物加帽 mRNA 翻译的蛋白质一定含有帽结合蛋白。分析结果显示 24kDa 蛋白自身很不稳定，但高分子质量复合物非常稳定。Sonenberg 与同事精确分析证明，纯化的活性复合物包含 3 个多肽：原来的 24kDa 帽子结合蛋白、分子质量为 50kDa 和 220kDa 的另外两个多肽（图 17.25）。现在已经给这些多肽重新命名，24kDa 帽子结合蛋白是 **eIF4E**，50kDa 多肽是 **eIF4A**，220kDa 多肽是 **eIF4G**。3-多肽复合物整体称作 **eIF4F**。

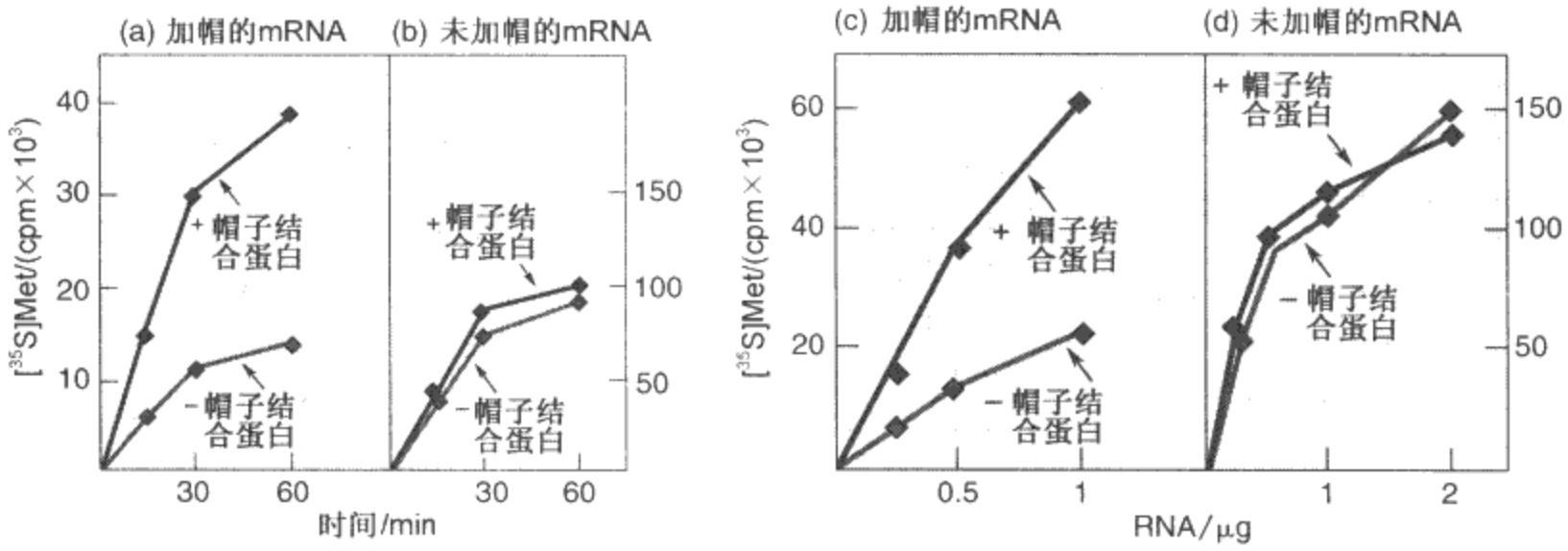


图 17.24　帽子结合蛋白促进有帽 mRNA 的翻译但不促进无帽 mRNA 的翻译。Shatkin 及同事在［^{35}S］甲硫氨酸存在时，用无 HeLa 细胞提取物研究有帽和无帽 mRNA 的翻译。图（a）和（c）有或无帽子结合蛋白时，有帽辛德毕斯病毒 mRNA 的翻译。（b）和（d）有或无帽子结合蛋白时，无帽的细小核糖核酸病毒（EMC）mRNA 的翻译。（*Source*：Adapted from Sonenberg, N., H. Trachsel, S. Hecht, and A. J. Shatkin, Differential stimulation of capped mRNA translation in vitro by cap-binding protein. *Nature* 285：331，1980.）

小结　eIF4F 是由 3 部分组成的帽子结合蛋白，eIF4E 具有实际的帽子结合活性，另两个亚基 eIF4A 和 eIF4G 协助其结合。

eIF4A 和 eIF4B 的功能　eIF4A 多肽是 eIF4F 的一个亚基，但它又能独立存在，具有独立功能，是 **DEAD 蛋白**家族的成员，该家族有共同氨基酸序列 Asp（D）、Glu（E）、Ala（A）、Asp（D）和 **RNA 解旋酶**（RNA helicase）活性，因此能打开经常在真核 mRNA 5′前导序列上发现的发夹结构。为有效发挥这一功能，eIF4A 需要 **eIF4B** 的协助。eIF4B 有一个 RNA 结合域，能促进 eIF4A 对 mRNA 的结合。Arnim Pause 和 Sonenberg 用一个很成熟的体外分析系统证明了 eIF4A 和 4B 的活性。他们从 eIF4A 和 4B 基因的细菌克隆产物开始，以避免其他真核蛋白的污染。然后加入标记的 RNA 解旋酶底物（图 17.26 左边）。该底物是具有 5′互补端的两个 40nt RNA 形成的 10bp RNA 双螺旋。如果 RNA 解旋酶解开此 10bp 结构，它就能分开这两个 40nt 的单体。通过电泳很容易区别单体和二聚体。单体形式越多，RNA 解旋酶活性越强。

如图 17.26 所示，少量的 eIF4A（有 ATP）引起微弱解旋（泳道 3），提示该因子有一定的自身 RNA 解旋酶活性。这种解旋酶活性受 eIF4B 激发（泳道 5），并依赖 ATP（比较泳道 4 和 5）。eIF4A 越多，产生的 RNA 解旋酶活性越高（泳道 6 和 7）。eIF4B 没有解旋酶活性（泳道 8）。因此，两个因子合作解开 RNA 螺旋及其发夹结构，这一活性依赖于 ATP。

小结　eIF4A 有 RNA 解旋酶活性，能够解开在真核生物 mRNA 5′前导序列中发现的发夹结构。这一功能由另一因子 eIF4B 协助，并需要 ATP 维持活性。

eIF4G 的功能　大多数真核生物 mRNA 是加帽的，帽子有助于核糖体结合。但有些病毒 mRNA 没

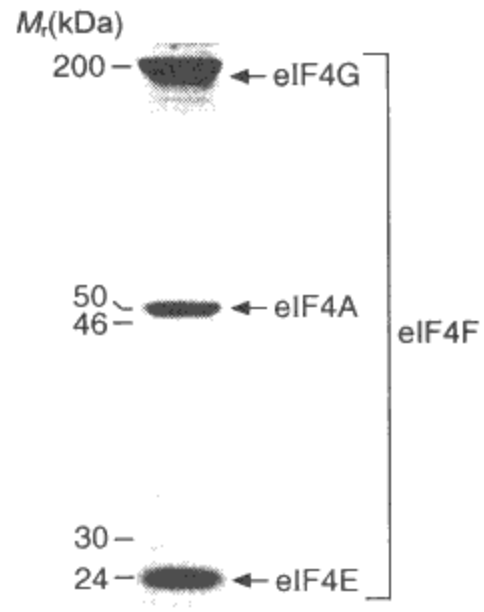


图 17.25　eIF4F（完整帽子结合蛋白）的组成。Sonenberg 及同事用一系列步骤包括 m^7GTP 亲和层析纯化了帽子结合蛋白，然后用 SDS-PAGE 显示纯化蛋白的亚基。亚基的分子质量和标准分子（200kDa、46kDa、30kDa）在左边标出。由 3 个多肽组成的复合物叫做 eIF4F。（*Source*：Edery，I.，M. Hümbelin，A. Darveau，K. A. W. Lee，S. Milburn，J. W. B. Hershey，H. Trachsel，and N. Sonenberg，Involvement of eukaryotic initiation factor 4A in the cap recognition process. *Journal of Biological Chemistry* 258（25 Sept 1983）p. 11400，f. 2. American Society for Biochemistry and Molecular Biology.）

有帽子，这类 mRNA 及少数细胞 mRNA 具有 IRES 来帮助核糖体结合。此外我们知道，mRNA 3′端的 poly（A）尾可促进翻译。至少在酵母细胞，后一过程涉及核糖体对 mRNA 的结合，这种结合需通过 poly（A）结合蛋白 **Pab1p**。eIF4G 蛋白参与所有这类翻译起始，作为与多种不同蛋白质相互作用的接头。

图 17.27 图示 eIF4G 参与翻译起始的 4 种不同途径。图 17.27a eIF4G 在普通的加帽 mRNA 上行使功能。eIF4G 的氨基末端结合 eIF4E，后者又结合到帽子上。eIF4G 的中央部分与 eIF3 结合，eIF3 再结合 40S 核糖体颗粒。因此，通过栓系 eIF4E 和 eIF3，eIF4G 使 40S 亚基靠近 mRNA 的 5′端，40S 亚基从那里开始扫描。图 17.27b 为细小核糖核酸病毒如脊髓灰质炎病毒对翻译起始的破坏。病毒蛋白酶切去 eIF4G 的氨基末端，使它不再与 eIF4E 作用识别帽子。因此，加帽的细胞 mRNA 不能被翻译。但是，eIF4G 的剩余部分仍然能结合脊髓灰质炎病毒 IRES 的 V 结构域，所以 40S 亚基仍被召集到病毒 mRNA 上。实际上，著名的能根除脊髓灰质炎的萨宾疫苗（Sabin vaccine）含有 3 个衰减的病毒株系。每个株系中重要的衰减事件是病毒 IRES 的 V 结构域发生改变，降低了对 eIF4E 的亲和性，从而有损病毒 mRNA 的翻译。图 17.27c 显示结合到 mRNA poly（A）尾的 Pab1p 如何与 eIF4G 相互作用，将 40S 亚基召集到 mRNA 的。最后，图 17.27d 显示同时发生的 eIF4G 和 eIF4E 的相互作用对帽子的结合与 eIF4G 和 Pablp 相互作用对 mRNA poly（A）尾的结合（译者注：图 17.27d 在第四版未给出）。eIF4G 在 mRNA 两端对蛋白质的双重结合有效地循

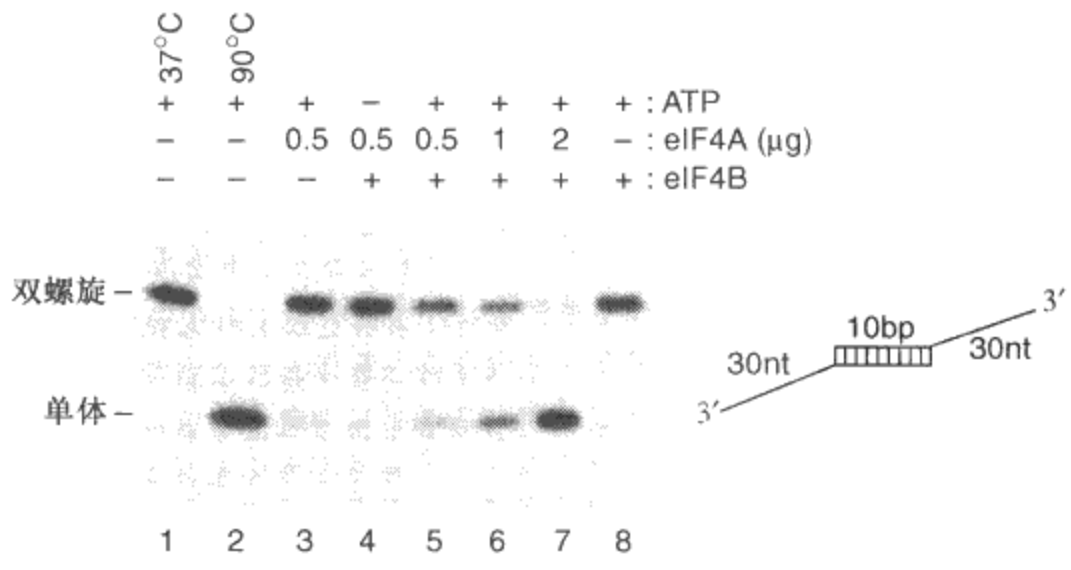


图 17.26　eIF4A 的 RNA 解旋酶活性。Pause 和 Sonenberg 检测了 ATP、eIF4A 和 eIF4B（如图上方指示）结合到放射性解旋酶底物上的情况（右边所示）。RNA 解旋酶解开 10bp 底物的双链区，将二聚物变成两个单体。二聚物和单体很容易通过凝胶电泳分离（如左边所示），用放射自显影检测。前 2 个泳道是在低、高温度时的底物，高温使底物的双链区溶解，产生单链。泳道 3～8 显示 ATP 和 eIF4A 是解旋酶活性所必需的，eIF4B 激发这一活性。（*Source*：Pause A. and N. Sonenberg，Mutational analysis of a DEAD box RNA helicase：The mammalian initiation translation factor elf-4A. *EMBO Journal* 11（1992）p. 2644，f. 1.）

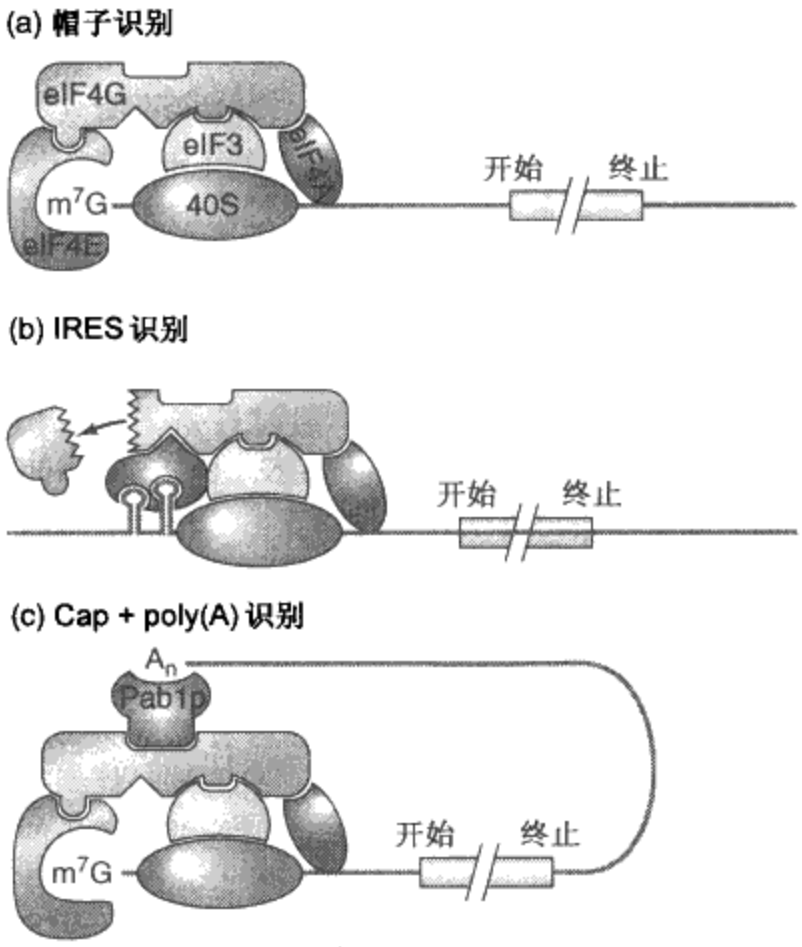


图 17.27　4 种不同情形下，eIF4G 在 40S 核糖体颗粒结合中的接头作用。（a）加帽的 mRNA。eIF4G 在 eIF4E 结合帽子和 eIF3 结合 40S 核糖体颗粒之间作为接头。这个分子链的形成将 40S 颗粒结合到 mRNA 上靠近帽子的位点，使其在此开始扫描。eIF4A 也结合到 eIF4G 上，但在这里所示的相互作用中不起作用。（b）具有 IRES 的 mRNA。IRES 可能与 RNA 结合蛋白（X）相互作用，X 再结合 eIF4G，确保对 40S 颗粒的召集。即使 eIF4G 的 N 端被细小核糖核酸病毒蛋白酶切除后，这种相互作用也会发生，细小核糖核酸病毒蛋白酶阻断对有帽细胞 mRNA 的结合。（c）poly（A）的参与。酵母 mRNA 3′端的 poly（A）结合 Pab1p，后者结合 eIF4G，因而有助于召集 40S 颗粒。（d）帽子结构和 poly（A）的协作。结合了帽子的 eIF4E 和结合了 poly（A）的 Pab1p 都与 eIF4G 结合，共同对 40S 颗粒的结合起作用。（*Source*：Adapted from Hentze，M. W.，elf4：A multipurpose ribosome adapter?*Science* 275：501，1997.）

环了 mRNA，从而至少在 3 方面影响翻译。第一，结合到 3′-UTR 的调节蛋白和 miRNA 靠近帽子，由此可帮助它们影响翻译起始；第二，完成一轮翻译的核糖体靠近帽子，可能促进再起始；最后，mRNA两端被隐藏起来，使降解 mRNA 的 RNase 不能接近。

小结　eIF4G 是一种适配器蛋白，与多种蛋白质结合，包括 eIF4E（帽子结合蛋白）、eIF3（40S 核糖体亚基结合蛋白）、Pab1p［poly（A）结合蛋白］。通过与这些蛋白质的相互作用，eIF4G 将 40S 核糖体亚基召集到 mRNA 上从而激活翻译起始。

eIF3 的结构和功能　我们刚才了解到 eIF3 与 eIF4E 和 eIF4G 合作将加帽的 mRNA 召集到 40S 核糖体颗粒上。我们也了解到它与 eIF4G 合作将带有 IRES 的细小核糖核酸病毒 mRNA 召集到 40S 核糖体颗粒上。我们还学习了 eIF3 阻止 60S 核糖体颗粒在起始过程完成前与 40S 颗粒结合。2005 年报道的 eIF3 结构研究有助于解释这些重要的行为。

Eva Nogales 及同事拍摄了 eIF3 和 eIF3 与其几种不同搭档结合的冰冻电子显微（cryo-electronic microscopy）照片，他们得到的结构分辨率很低（约 30Å），但是通过比较有无搭档时的结构，可以清楚地看出搭档相对于 eIF3 的位置。首先，图 17.28a 显示人类 eIF3 自身 3 个不同方向的结构，它是个复杂的蛋白质，至少含有 12 个多肽，其结构也相应地复杂，像人形，有头、两个胳膊、两条腿。

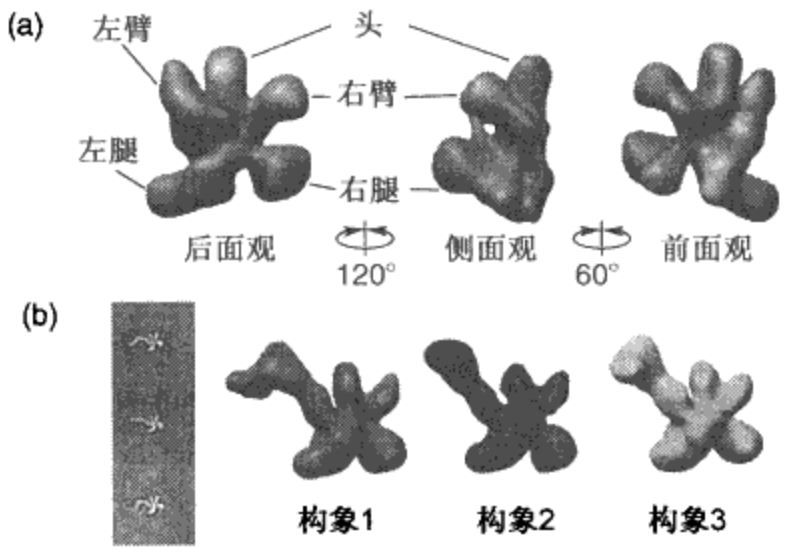


图 17.28　人类 eIF3 和 eIF-HCV IRES 复合物的冰冻电镜（cryo-EM）结构。（a）eIF3 在 3 个方向的结构。不同视角间的旋转角度在底部标出；（b）eIF-IRES 复合物的 3 种不同构象结构。左边的电镜图显示这 3 种构象的平均外貌。彩色 3D 模型源自这 3 种平均外貌。构象 1 占优势。（*Source*：Reprinted with permission from *Science*，Vol. 310，Bunpote Siridechadilok，Christopher S. Fraser，Richard J. Hall，Jennifer A. Doudna，and Eva Nogales，"Structural Roles for Human Translation Factor elf3 in Initiation of Protein Synthesis，" Fig 1，p. 1513，Copyright 2005，AAAS.）

接着他们得到了 3 种不同形式的丙型肝炎病毒（HCV）未加帽 mRNA 的 IRES 与 eIF3 复合物的结构。与细小核糖核酸病毒中的 IRES 通过 eIF4G 与 eIF3 相互作用不同，HCV 的 IRES 可直接与 eIF3 相

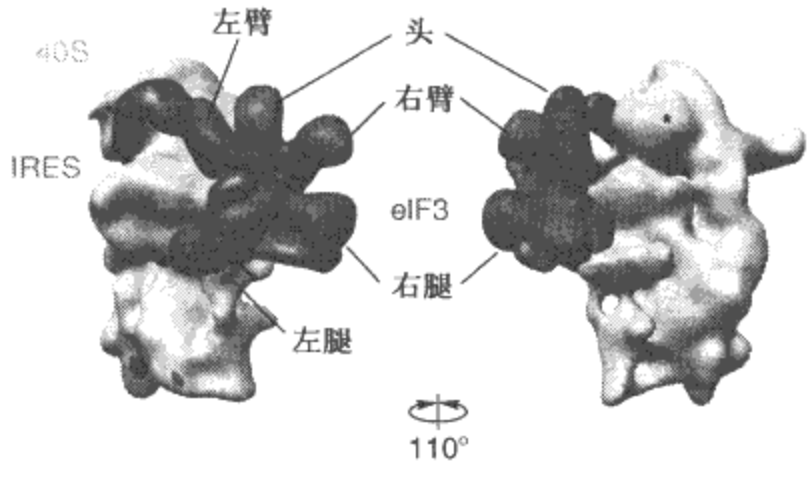


图 17.29　eIF3-IRES-40S 复合物模型。两种视图的差别在于标出的旋转角度。复合物 3 部分的颜色分别为：eIF3 洋红色，IRES 紫色，40S 核糖体颗粒黄绿色。星号标出核糖体结合的 C 激酶（RACK）的位置。（*Source*：Reprinted with permission from *Science*, Vol. 310, Bunpote Siridechadilok, Christopher S. Fraser, Richard J. Hall, Jennifer A. Doudna, and Eva Nogales, "Structural Roles for Human Translation Factor eIF3 in Initiation of Protein Synthesis," Fig. 3a, p. 1514, Copyright 2005, AAAS.）

互作用，因此替代了 eIF4E 和 eIF4G。图 17.28b 显示出在 eIF3 左臂的肩上明显多出一个电子密度区，可能与 IRES 对应，但是仔细比较复合物的最普通形式（构型-1）与 eIF3 自身的结构，可发现 eIF3-IRES 复合物的复合视图。

在前期的 cryo-EM 研究中，建立了 IRES-40S 核糖体颗粒模型，因此 Nogales 及同事可以将他们的 eIF3-IRES 模型与 IRES-40S 模型放在一起，建立 eIF3-IRES-40S 颗粒复合物（图 17.29）。在该复合物中，eIF3 的位置有助于解释 eIF3 阻止 40S 和 60S 核糖体颗粒过早结合的活性，eIF3 位于两个颗粒之间的关键接触点上，因此可以阻挡它们相互接触。

接下来，Nogales 及同事比较了 eIF3-eIF4G 与 eIF3 的结构，确定 eIF4G 相对于 eIF3 的位置。如图 17.30a 所示，eIF4G 结合到 eIF3 的左臂尖上与 IRES 的突出部分相同的位置。这有助于解释 IRES 为什么能替代 eIF4G，这两个分子在 eIF3 上有相同的作用位点，因此推测两者以相似的方式召集 eIF3 和 40S 核糖体颗粒。

知道了 eIF4G 在哪儿结合 eIF3，Nogales 及同事就可以推断在 eIF4G-IRES-40S 颗粒复合物中 eIF4G 的位置了。图 17.30b 给出了这个复合物的模型，包括 mRNA 结合到该核糖体颗粒上的路径。eIF4G 的位置靠近 mRNA 5′端的帽子。这很有意思，因为帽子结合蛋白 eIF4E 是结合 eIF4G 的，它应该在与帽子直接作用的位置。

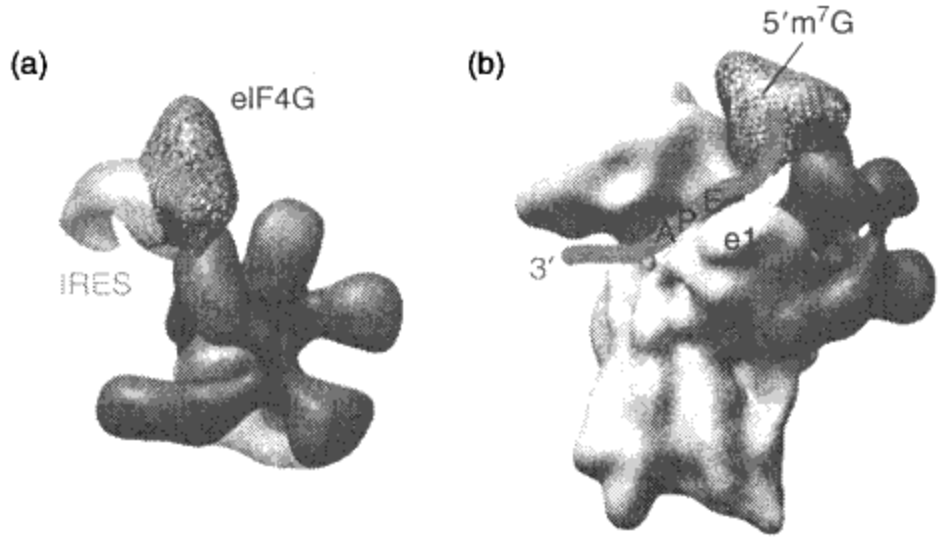


图 17.30　图谱定位 eIF4G 在 eIF3 上的结合位点和 IF3-eIF4G-40S 复合物模型。(a) eIF3-IRES 模型，特别强调 eIF4G 的位置，注意在 eIF4G 左下角 IRES 和 eIF4G 的重叠；(b) IF3-eIF4G-40S 复合物模型，标出了 mRNA 5′端的路径，A、P、E 是 tRNA 结合在核糖体上的位点，将在第 18 章和第 19 章讨论。5′m^{7}G 指 mRNA 帽子，e1 指 eIF1 的位置，颜色标注同图 17.29。（*Source*：Reprinted with permission from *Science*, Vol. 310, Bunpote Siridechadilok, Christopher S. Fraser, Richard J. Hall, Jennifer A. Doudna, and Eva Nogales, "Structural Roles for Human Translation Factor eIF3 in Initiation of Protein Synthesis," Fig. 4 a&c, p. 1515, Copyright 2005, AAAS.）

小结　Cryo-EM 研究表明，eIF3 是个 5 圆凸的蛋白质，可以通过相同位点结合 eIF4G 和病毒 IRES 的突出部分。这个发现揭示了 IRES 如何替代 eIF4G 召集 eIF3，进而召集 40S 核糖体亚基到 mRNA 上。这些研究产生了 eIF3-IRES-40S 复合物模型，解释了 eIF3 如何阻止 40S 和 60S 的过早结合，eIF3 阻断了这两个核糖体颗粒间的关键接触点。研究结果还表明 eIF4G 进而是 eIF4E 被定位到靠近结合在 40S 核糖体颗粒上的 mRNA 帽子处，所以 eIF4E 处在执行其帽子结合任务的位置上。

eIF1 和 eIF1A 的功能 在体外 eIF1 对翻译活性只起很小的激发作用（约 20%），因此，长期以来都认为 eIF1 的作用可有可无。然而，编码 eIF1 和 eIF1A 的基因对酵母活力至关重要，所以它们的产物不会可有可无。那么，它们起什么作用呢？Tatyana Pestova 及同事发现了答案。没有 eIF1 和 eIF1A，40S 亚基只能扫描几个核苷酸，如果根本没有 eIF1 和 eIF1A，它只能松散地结合 mRNA。有了这些因子，40S 颗粒才能扫描起始密码子，形成稳定的 48S 复合物。

Pestova 及同事利用基于引物延伸技术（第 5 章）的**趾纹分析**（toeprint assay）定位 40S 核糖体亚基结合 mRNA 时的前沿（leading edge）。他们分离出 40S 亚基和哺乳动物 β-球蛋白 mRNA 的复合物，然后将复合物与引物混合，该引物与 mRNA 起始密码子的下游结合。用核苷酸和反转录酶延伸引物，当反转录酶碰上 40S 亚基的前沿时就停下来，所以延伸引物的长度可表明前沿的位置。如果将 40S 亚基看成一只脚，其前沿就是脚指头，这就是为什么我们称其为趾纹分析。最后，Pestova 及同事对引物延伸产物做电泳分析，测定其大小。图 17.31 给出了这个过程的示意图。

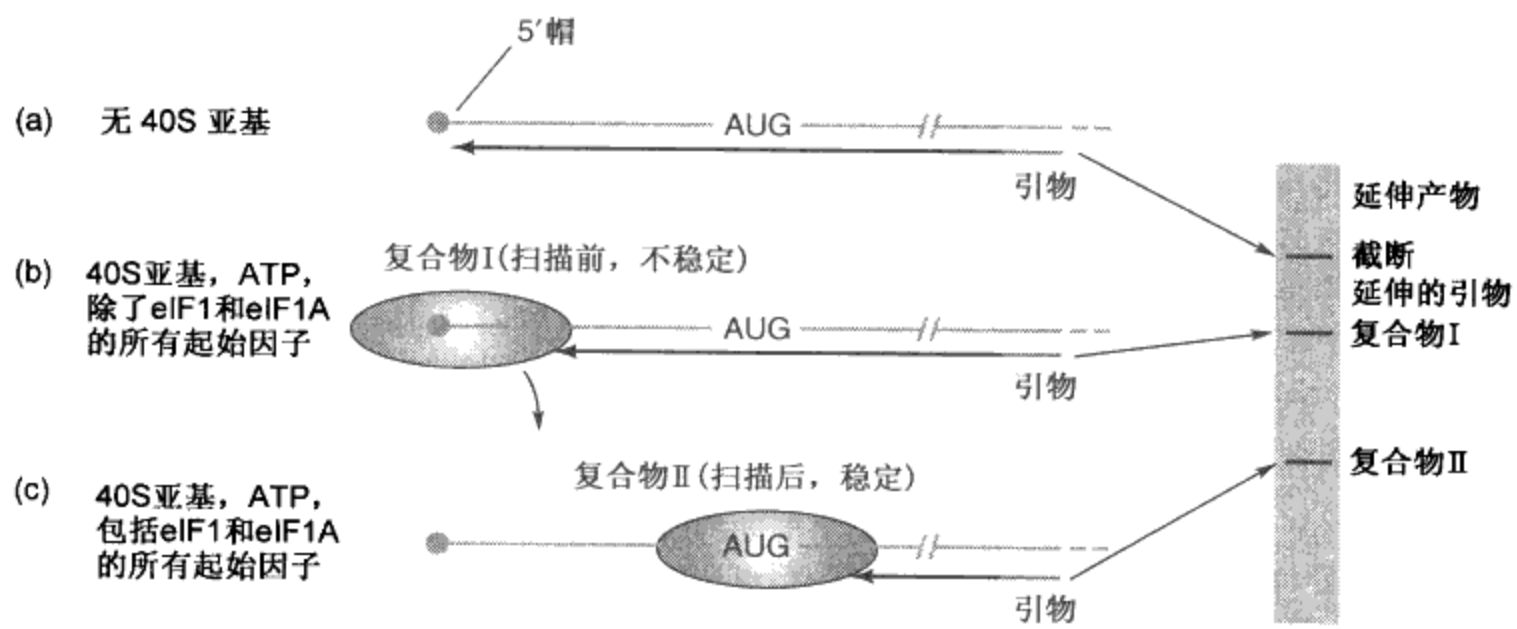


图 17.31 趾纹分析原理。（a）阴性对照。省去重要成分 40S，所以 40S 核糖体亚基与 mRNA 不能形成复合物。没有 40S 颗粒阻断反转录，引物可延伸至 mRNA 的 5′端，产生相应于裸 mRNA 的截断延伸的引物。（b）缺乏 eIF1 和 eIF1A 时形成的复合物。加入左边列出的所有成分，但省去 eIF1 和 eIF1A。复合物 I 在帽子处形成，但是不再前行。因此，引物延伸了很长距离到达 40S 颗粒的前沿。（c）在 eIF1 和 eIF1A 存在时形成的复合物。40S 颗粒扫描下游，在起始密码子（AUG）处形成稳定复合物（复合物 II）。因此，引物在被 48S 复合物中 40S 颗粒的前沿挡住之前只延伸了一段很短的距离。（*Source*：Adapted from Jackson，R. J.，Cinderella factors have a ball. *Nature* 394：830，1998.）

实际结果见图 17.32。泳道 1 和 2 只含有 mRNA或 mRNA 与 40S 亚基，没有起始因子，所以没有复合物形成。泳道 3 含有 mRNA、40S 亚基、eIF2、eIF3、eIF4A、eIF4B 和 eIF4F。这些因子只促进复合物 I 的形成，没有复合物 II 的痕迹。在这些条件下，40S 颗粒的前沿相对于 mRNA 的帽子在＋21 到＋24 之间。如果 40S 亚基结合到帽子上尚未开始扫描或只扫描了最短距离，这是在我们预计的位置附近。泳道 4 含有泳道 3 的所有因子，再加上起始因子的混合物，起始因子是用缓冲液洗涤核糖体，然后收集可被浓度为 50%～70% 硫酸铵沉淀的蛋白质而获得。很明显，起始因子混合物加上其他因素可促进复合物 II 的形成，其前沿相对于起始密码子 AUG 的 A 在＋15 和＋17 之间。如果 40S 颗粒位于起始密码子的中心，这基本上是我们所预计的位置。

接下来，Pestova 及同事纯化了 50%～70% 硫酸铵组分中的重要蛋白质，匀浆获得部分氨基酸序列，鉴定结果证明是 eIF1 和 eIF1A。图 17.32 泳道 5 和 6 表明，这两个因子单独都不能激发复合物 II 的形成。泳道 7 证明它们合起来可专一性地引起复合物 II 形成。因此，这两个因子相互配合促进复合物 II 的形成。在泳道 8，先让复合物 I 形成 5min，然后加入 eIF1 和 eIF1A，在这种条件下只有复合物 II 形成。因此，复合物 I 并不是最终产物，起始因子可以将其转换为复合物 II。

eIF1 和 eIF1A 将复合物 I 转换为复合物 II 只是简单地引起 40S 亚基在同一 mRNA 上扫描到更远处，还是这两个因子引起 40S 颗粒与 mRNA 解离然后再扫描起始密码子呢？为了找出答案，Pestova 及同事在放射性标记的 mRNA 上形成复合物 I，然后在有或无 15 倍过量未标记竞争 mRNA 条件下，加

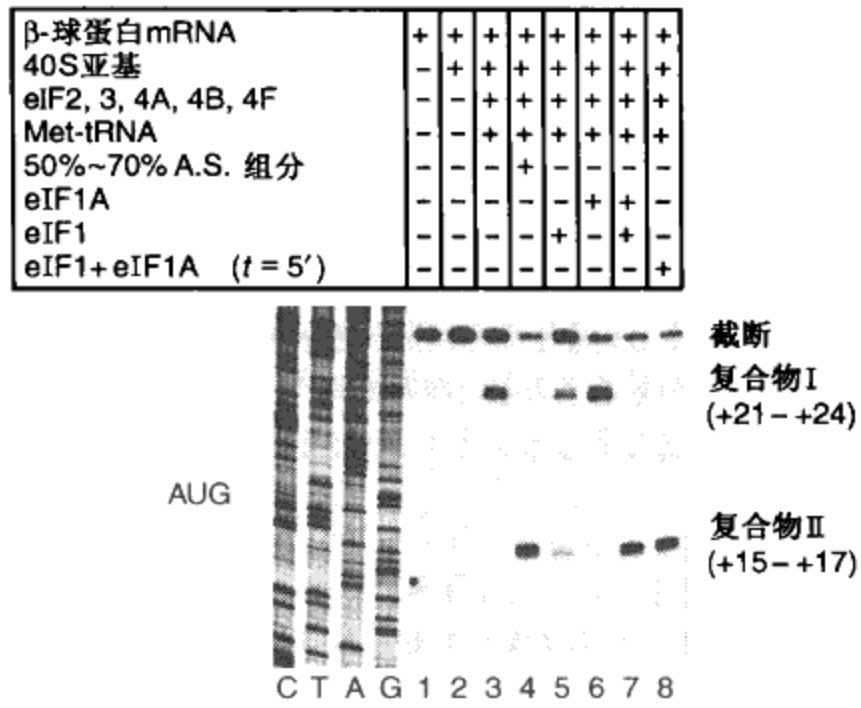


图 17.32　趾纹分析结果。Pestova 及同事按图 17.31 所示以哺乳动物的 β-球蛋白 mRNA 做趾纹分析实验。每个分析所加入的成分在泳道 1～8 的上部列出。“50%～70%A.S. 组分”（泳道 4）是指通过离心沉淀 50%～70%浓度硫酸铵核糖体洗涤液中的蛋白质而获得的起始因子。“eIF1＋eIF1A（t=5′）”是指加入其他因子 5min 之后再加 eIF1 和 eIF1A。泳道 C、T、A 和 G 是对相应于 β-球蛋白 mRNA 的 DNA 测序结果，放上这些测序泳道是作为标准来确定复合物中 40S 核糖体颗粒的确定前沿（趾纹）的准确位置。起始密码子（AUG）的位置在左边给出，相应于全长流失延伸引物、复合物 I、复合物 II 的条带在右边给出，包括 40S 颗粒相对于帽子和起始密码子的前沿。eIF1 和 eIF1A 为复合物 II 形成所必需。（*Source*：Pestova，T. V.，S. I. Borukhov，and C. V. T. Hellen，Eukaryotic ribosomes require initiation factors 1 and 1A to locate initiation codons. *Nature* 394（27 Aug 1998）f. 2，p. 855. Copyright © Macmillan Magazines Ltd.）

入 eIF1 和 eIF1A，用蔗糖梯度超速离心纯化 48S 复合物（预计是复合物 II 的等价物），然后用液体闪烁计数（第 5 章）测定复合物的放射性。

正如所预期的（图 17.33），他们发现在没有竞争 mRNA 时有一个清楚的 48S 放射性峰，但是在温育开始就加入竞争 mRNA 时，或者在复合物 I 形成 5min 后加入竞争 mRNA 时，没有出现 48S 的放射性峰。因此，eIF1 和 eIF1A 不是简单地使 40S 亚基扫描下游并在相同的标记 mRNA 上形成复合物 II，否则在复合物 I 形成 5min 后加入这些因子和竞争 mRNA时，应当能看见标记的 48S 复合物。相反，这些因子干扰了在标记 mRNA 上的复合物 I，迫使在过量未标记 mRNA 上形成新的复合物。预计 40S 亚基放弃了标记的 mRNA，结合到未标记 mRNA 的帽子上，扫描其起始密码子，形成复合物 II。因此，eIF1 和 eIF1A 不仅对适当 48S 复合物形成很重要，它们也破坏了 40S 亚基与 mRNA 之间的不适当复合物。

小结　eIF1 和 eIF1A 一起促使稳定 48S 复合物的形成，该复合物包括起始因子、Met-tRNA$_i^{Met}$及结合在 mRNA 起始密码子上的 40S 核糖体亚基。eIF1 和 eIF1A 表现为解离 40S 核糖体亚基与 mRNA 之间不合适的复合物，促进 48S 稳定复合物的形成。

eIF5 和 eIF5B 的功能　一旦 eIF2 将 Met-tRNA 送到 40S 核糖体亚基上，并且 mRNA 也结合上来形成完整 48S 起始复合物之后，eIF2 需从复合物上解

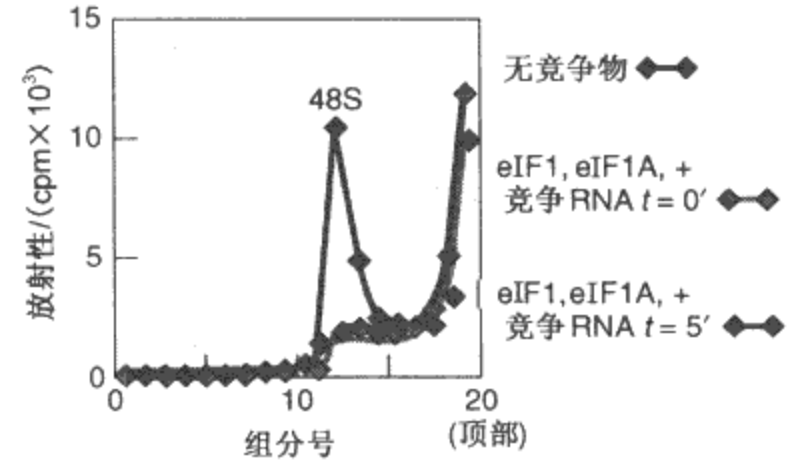


图 17.33　竞争 RNA 对 48S 复合物形成的影响。Pestova 及同事将 [^{32}P]-β-球蛋白 mRNA 与 40S 核糖体颗粒加上起始因子和未标记的竞争 RNA 组合共孵育，组合如右边所示：无竞争 RNA（蓝色）；竞争 RNA（绿色），在零时加入 eIF1 和 eIF1A；竞争 RNA（红色），在孵育 5min 时（复合物 I 已形成）加入 eIF1 和 eIF1A。孵育之后，对混合物进行蔗糖梯度超速离心，测定稳定的 48S 复合物形成，包括 40S 颗粒、[^{32}P]mRNA、Met-tRNA$_i^{Met}$。用组分数与每组分液体闪烁计数的放射性（每分钟次数，cpm）作图。梯度的上部在 19 号，如图右下角所示。（*Source*：Adapted from Pestova，T. V.，S. I. Borukhov，and C. V. T. Hellen，Eukaryotic ribosomes require initiation factors 1 and 1A to locate initiation codons. *Nature* 394：856，1998.）

离下来。完成这一过程需要 GTP 水解，但与 IF2 不同的是，eIF2 需要另一个因子 **eIF5** 来帮助自己水解

所结合的GTP。即使在eIF5诱导结合在eIF2上的GTP水解之后，48S复合物仍然不能接受60S核糖体亚基来完成起始过程，还需要另外一个因子**eIF5B**的参与。

2000年，Christopher Hellen及同事在检测eIF2解离之后，eIF5诱导60S核糖体亚基结合48S复合物的能力时发现了eIF5B。他们发现单独eIF5不足以引起核糖体亚基的结合，但是高离子强度缓冲液洗涤核糖体所得蛋白质复合物能够与eIF5一起引起核糖体亚基的结合。他们从这种“盐洗脱”物中，纯化出eIF5B，该蛋白质有诱导-接合活性。单独纯化的eIF5B（或通过基因克隆而得到的修饰过的eIF5B）不能诱导亚基结合，但是在其他因子存在时它能促使亚基的结合，这些因子包括eIF1、eIF2、eIF3和eIF5。

那么，这种亚基-接合反应需要GTP水解吗？将事先形成的48S复合物与eIF5、eIF5B、60S亚基以及GTP或其不可水解的类似物——GDPNP分别混合。没有GTP或GDPNP时，无亚基结合发生。因此，GTP是需要的。并且，GDPNP能够支持亚基结合，但需要化学平衡量的eIF5B。此外，eIF5B在激发亚基接合过程中与GTP一起发挥催化作用。因为GDPNP足以支持结合，所以GTP水解对亚基结合是非必需的。

Hellen及同事也发现GDPNP存在时形成的eIF5B不能从80S复合物上释放，但GTP存在时形成的eIF5B能从复合物上释放。因此，GTP水解是eIF5B从核糖体释放所必需的。在这方面，eIF5B与原核生物的IF2类似，IF2从核糖体释放需要GTP水解。这两个因子在具有核糖体激发的GTPase方面也相似，它们都在核糖体亚基接合方面起相似的作用。事实上，一方面，这两个因子是同源的，所以它们功能相似不足为奇；另一方面，eIF5B又与IF2完全不同，它不能促进Met-tRNA$_i^{Met}$的结合，而IF2却能在原核生物中执行这一功能。在真核生物中是eIF2有这样的功能，而不是eIF5B。

小结 eIF5B与原核生物的IF2同源。eIF5B在结合GTP和激发两个核糖体亚基的结合方面与IF2相似。在这个反应中，eIF5B与eIF5共同发挥作用。eIF5B在利用GTP的水解促进自身与核糖体分离，使蛋白质合成能够开始方面与IF2也很相似，与IF2不同之处在于它不能促使氨酰-tRNA结合到核糖体小亚基上。在真核细胞中这一任务由eIF2完成。

17.3 翻译起始的调控

我们已经研究了真核细胞基因在转录及转录后水平的表达调控，但调控也在翻译水平发生。既然已具有转录水平的广泛调控，有机体为什么还要进化出翻译水平对基因表达的调控机制呢？翻译调控的最大优势是速度。只需启动已有mRNA的翻译，新基因产物就能够很快地产生，这对真核生物尤其重要，因为其转录物相对较长，需要较长的时间去合成。很自然，大多数的翻译调控发生在起始阶段。

细菌的翻译调控

我们已经学过多数细菌基因表达的调控发生在转录水平。绝大多数细菌mRNA非常短的寿命（只有1～3min）与这一机制相符，这样能使细菌迅速响应变化的环境。虽然多顺反子转录物上的不同顺反子被翻译的程度不同，例如，*LacZ*、*Y*和*A*顺反子产生蛋白质的摩尔比是10∶5∶2，但这个比率在多数情况下是常量，它仅反映出3个顺反子的Shine-Dalgarno序列的相对效率和多顺反子mRNA各部分降解速度的不同。但确实有真正的细菌翻译调控的例子存在，我们来讨论一下其中的几个。

mRNA二级结构的转换 从本章的图17.9可见，mRNA的二级结构在翻译效率方面起作用。RNA噬菌体MS2家族的复制酶顺反子的起始密码子被埋在双链结构中（它们是外壳基因的一部分），由此解释了为什么噬菌体的复制酶基因须在外壳蛋白翻译后才能被翻译：核糖体沿着外壳基因移动，打开藏有复制酶基因起始密码子的二级结构。

另一个通过mRNA结构调控翻译的例子是第8章提到的在*E. coli*热激过程中σ^{32}合成的诱导。当*E. coli*细胞经历从正常的37℃到42℃的温度变化时，会打开一套热激基因来应对温度升高。这些新的热激基因对σ^{32}而不是正常的σ^{70}做出响应。但是，σ^{32}在热激后不到1min就开始累积，这段时间对于σ^{32}基因（*rpoH*）的转录和相应mRNA的翻译来说太短了，怎样解释σ^{32}的这种迅速累积呢？

有两种答案。第一，先前存在的通常是不稳定的σ^{32}变得稳定了；第二，与我们在这里讨论的更相关的是σ^{32}基因在翻译起始水平受到调控。编码σ^{32}的mRNA通常折叠成其起始密码子被隐藏在二级结构中的形式。也就是说，起始密码子与mRNA下游区域的其他碱基配对。当温度升高时，碱基配对引起的二级结构变性，暴露出起始密码子，mRNA被翻译。因此这个特殊σ因子总是有许多mRNA，但是在温度升到危险水平之前是不能被翻译的。换句话说，这种在mRNA上的内在热感受器允许加热来激发在翻译水平的基因表达。

1999年，Takashi Yura及同事用*rpoH*基因的

衍生物提供了强有力的证据支持这一假说，*rpoH*′基因产生如图 17.34 所示的二级结构 mRNA，衍生物 mRNA 与野生型 mRNA 具有相同的调控特征。注意，在起始密码子和 mRNA 靠近 3′端区域的碱基配对形成“茎 I”可阻止该 mRNA 在生理条件下的翻译。接着，Yura 及同事在茎 I 处进行突变，使碱基配对更强或更弱，然后测定这些突变对热诱导的效应。

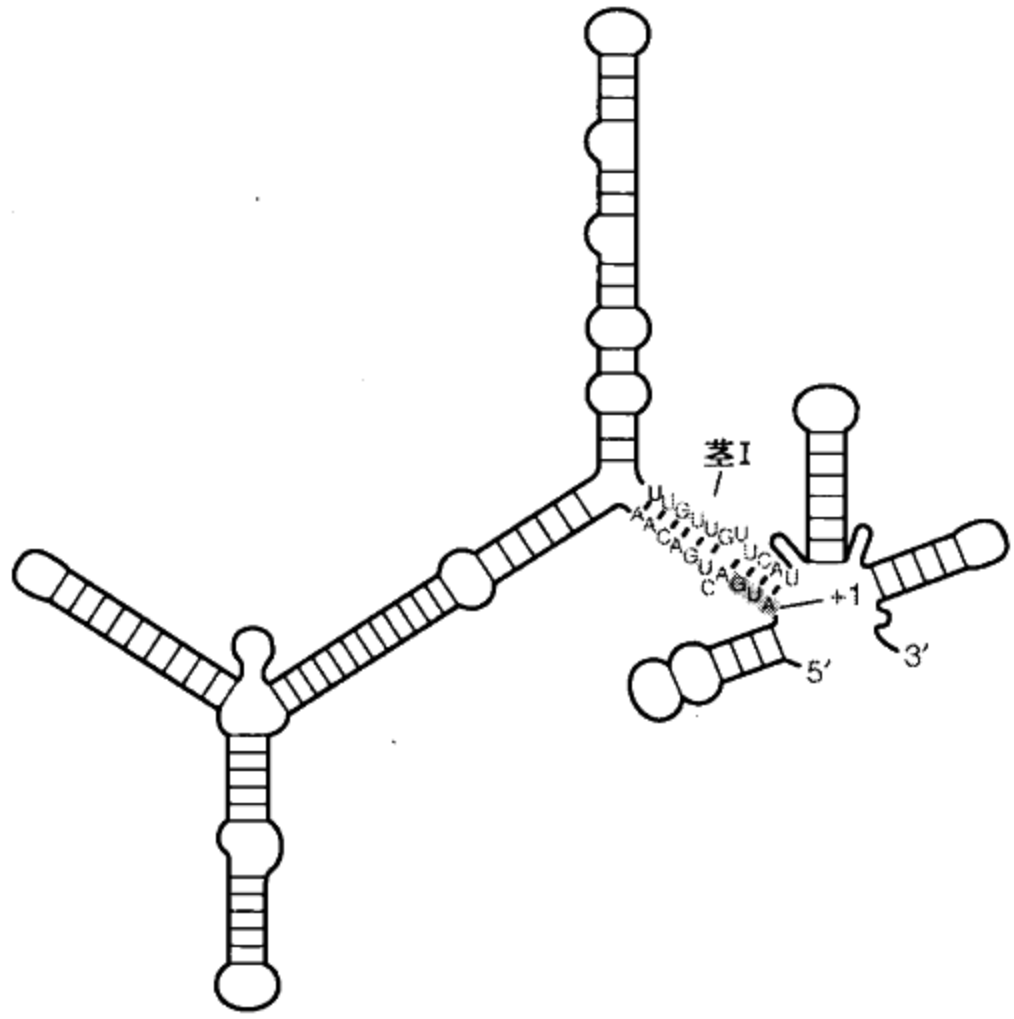


图 17.34　部分 *rpo*H mRNA 的二级结构。显示茎 I 碱基配对区的序列，包括起始密码子 AUG（加框的）。（*Source*：Adapted from Morita，M. T.，Y. Tanaka，T. S. Kodama，Y. Kyogoku，K. Yanagi，and T. Yura，Translational induction of heat shock transcription Factor σ^{32}. Evidence for a built-in RNA thermosensor. *Genes and Development* 13 [1999] p. 656，f. 1b.）

当突变使茎 I 的碱基配对更强时，诱导作用被减弱。例如，相对于 AUG 密码子 A 的＋5 位的 C，在正常情况下与对应链的 U 不配对，但是当这个 C 变为 A 时，就可以与 U 配对，使茎 I 的稳定性增加 2.9kcal/mol，由此将诱导作用由正常的 3.5 倍降低到只有 1.4 倍。这一结果是合理的，因为强的碱基配对更难被热打断。此外，减弱碱基配对的突变在高温和低温下都能增强基因的表达。这也是合理的，因为较弱的碱基配对即使在较低的温度下也容易被打断。

小结　细菌 mRNA 非常短命这一事实意味着转录调控在这类生物中是一种非常有效的控制基因表达的方式，但是翻译调控也是存在的。mRNA 二级结构可以控制翻译的起始，例如，在 MS2 类噬菌体中，起始密码子埋在二级结构中直到核糖体翻译衣壳基因后才打开这个结构。在另一个例子中，*E. coli* 热激蛋白 sigma 因子 σ^{32} mRNA的起始密码子受二级结构抑制，该二级结构可被加热破坏，因此，热激可引起 σ^{32} mRNA 起始密码子的立刻暴露，暴发 σ^{32} 迅速合成。

被蛋白质和 RNA 诱导的 mRNA 二级结构的转换　在第 16 章我们学过叫做 microRNA 的小 RNA 可以控制真核生物 mRNA 的稳定性和翻译。细菌中的翻译也能被一类叫做**小 RNA**（small RNA，**sRNA**）的短 RNA 所控制，它们作用于 mRNA 的二级结构。例如，胁迫 sigma 因子（*rpoS*）mRNA 的起始密码子通常是埋在二级结构中的，几乎没有蛋白质合成。但是，如图 17.35 所示，DsrA sRNA 与伴侣蛋白 Hfq 合作可以与 mRNA 的上游区域碱基配对，暴露 *rpoS* 的起始密码子使翻译发生。

我们在第 7 章学过，**核开关**（riboswitch）是 mRNA 的内部区域，可因结合小分子而改变构象，从而开启或关闭基因表达。例如，通过从衰减子到终止子的转换，引起翻译的衰减。RNA 里结合小分子的区域叫做**核酸适体**（aptamer）。

Ronald Breaker 及同事在 2002 年发现的核开关是最早发现的例子之一。他们发现 *E. coli* 编码合成硫胺（维生素 B_1）所需酶类的 mRNA 至少有 2 种构象。当硫胺或硫胺焦磷酸结合到 mRNA 的一个核酸适体时，mRNA 的构象隐藏核糖体结合位点，所以 mRNA 不能翻译。这当然是有益的，因为硫胺的存在指示细胞不要浪费能量合成更多的合成硫胺的酶。注意，在这个核开关中没有蛋白质介入，小分子硫胺自己就能改变 mRNA 的构象。

Breaker 及同事也证明合成辅酶 B12 的酶中，编码其中一个酶的 mRNA 的前导序列（leader）可以

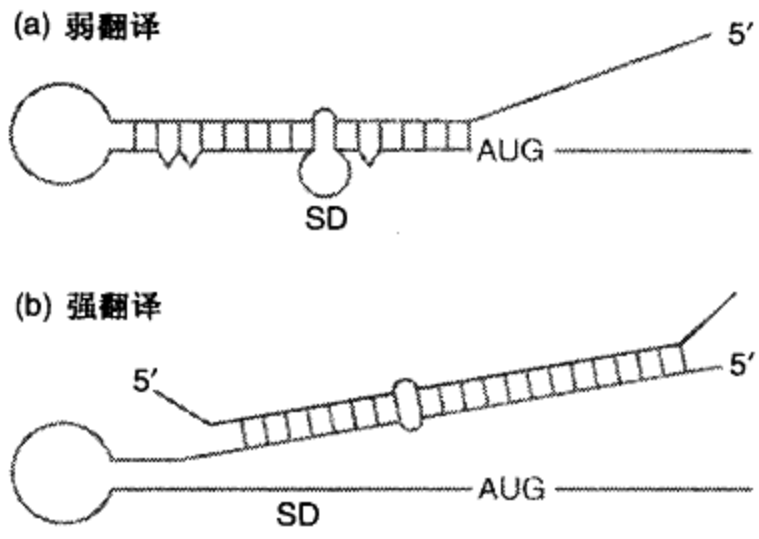


图 17.35　sRNA 激活的 *rpoS* mRNA 翻译的模型。（a）*rpoS* mRNA 的 5′-UTR 内的碱基配对产生茎环，隐藏了 SD 序列和起始密码子（AUG）。（b）DsrA sRNA 结合到 RNA 结合蛋白 Hfq 上并与部分 5′-UTR 碱基配对，打开 SD 序列和起始密码子以便与核糖体结合。

结合到辅酶上，引起该 mRNA 结构的改变，这个结构对调控辅酶的合成非常重要。他们猜想是否有一种共同的机制适用于硫胺的生物合成途径，因为编码该途径酶类的其中 2 个基因（*thiM* 和 *thiC*）都含有 *thi box* 保守的序列和二级结构。

相应地，他们将两种 *thi* 盒子连接到 *lacZ* 报告基因上，然后检测这个重组子在有或无硫胺条件下产生 β-半乳糖苷酶的能力，发现硫胺 β-半乳糖苷酶的合成分别下降了 18 和 110 倍。因此，*thi* 盒子确实参与基因活性的抑制。在 *thiC* 重组子，*thi* 的抑制作用大多数都是转录水平的，而在 *thiM* 重组子，*thi* 的抑制作用全部是翻译水平的。由于本章是学习翻译调控，因此我们集中讨论 *thiM* 基因。

接着，Breaker 及同事应用**串联探测**（in-line probing）技术检测硫胺或其衍生物是否引起 mRNA 前导序列的结构变化。该策略是基于无结构 RNA 对自发切割的敏感性高于有很多二级结构（分子内的碱基配对）或三级结构（三维结构）的 RNA。所以研究者将 *thiM* mRNA 在有或无硫胺素焦磷酸（TPP）条件下孵育 40h，然后对产物电泳，检测在何处发生了切割。图 17.36a 揭示，在两种情况下都有大量切割发生，但是有显著区别。具体讲，在 TPP 存在下，区间 39～80（包括 *thi* 盒）很少发生切割。

还要注意星号所指区间（126～130），这是在 TPP 存在下唯一较有序的区域（很少切割），除了 *thi* 盒及其 5′侧（5′-side）的几个核苷酸。这个区域涵盖核糖体结合的 Shine-Dalgarno 序列。这些结果提示 TPP 引起 *thiM* mRNA 构象的转换，使 Shine-Dalgarno 序列隐藏在碱基配对茎中，由此阻止核糖体结合，降低 mRNA 翻译的效率。

Breaker 及同事鉴定出 GAAG 序列，见图 17.36b 中橙色加亮区。GAAG 紧接在 *thi* 盒的后面，可以与 P8 茎的 Shine-Dalgarno 序列对面的 108～111 位（也用橙色加亮）的 CUUC 碱基配对。由此提示 CUUC（108～111 位）碱基在正常情况下与 *thi* 盒末端的 GAAG 碱基配对，使 Shine-Dalgarno 序列为核糖体所利用。这种 mRNA 结构允许进行活跃的翻译，而 TPP 通过与 *thi* 盒的核酸适体结合，改变了 mRNA 的结构，使 108～111 位的 CUUC 与 Shine-Dalgarno 序列的 GGAG 碱基配对，隐藏了 Shine-Dalgarno 序列，核糖体不能结合，从而减慢翻译。

由这个假设可以做出几个预测。第一，含有 *thi* 盒的一段 mRNA 应当对低浓度的 TPP 做出响应。Breaker 及同事确实发现 165nt 长、含有 *thi* 盒（165*thiM* RNA）mRNA 的结构性修饰在只有 600nmol/L 的浓度时是半完整的（half-complete）。第二，TPP 应当紧密地结合到 165*thiM* RNA 上，Breaker 及同事利用平衡透析技术证明它确实是紧密结合的。平衡透析利用标记的配体（本例中氚-标记的 TPP）和另一个大分子（165*thiM* RNA），将它们分别放在两个小室中，隔以透析膜，使小分子的 TPP 可以透过，但大分子的 RNA 被滞留。当两个室之间建立了平衡后，实验人员测定了每个室中的标记量，由此得出解离常数。本例中，含 RNA 小室的标记量远远高于另一个室，表明是解离常数很低的（在 TPP 和 RNA 之间有紧密结合）。

第三个预测是硫胺家族成员与 *thiM* RNA 的结合应当是专一性的。确实，硫胺、硫胺磷酸、TPP 与 RNA 结合得很好，但是氧化硫胺及其他衍生物结合得不好。最后，对 RNA 做修饰干扰了 *thiM* 前导序列的重要结构元件，会阻断 TPP 的结合和 *thiM* 表达的控制。Breaker 及同事通过对参与推测茎 P3、P5、P8 的一些碱基的修饰验证了这个预期结果。突变 RNA 都不能与 TPP 结合，在 TPP 存在下没有表现出 *thiM* 表达的降低。但是恢复 P3、P5、P8 茎碱基配对的补偿突变都恢复了 TPP 结合和 *thiM* 控制。例如，将 106 位碱基由 U 变为 G，107 位碱基由 A 变为 C，可阻断 130 和 131 位分别与 A 和 C 的碱基配对，从而使 P8 减弱，阻断了 TPP 的结合和控制。但是，如果 130 和 131 位的 A 和 C 分别被变为 G 和 U，TPP 的结合和控制得到恢复。因此，在这 3 个茎中的碱基配对看上去对控制都很重要，正如这个假说所预测的那样。

小结　小 RNA 与蛋白质协调作用可影响 mRNA 二级结构控制翻译的起始。核开关通过

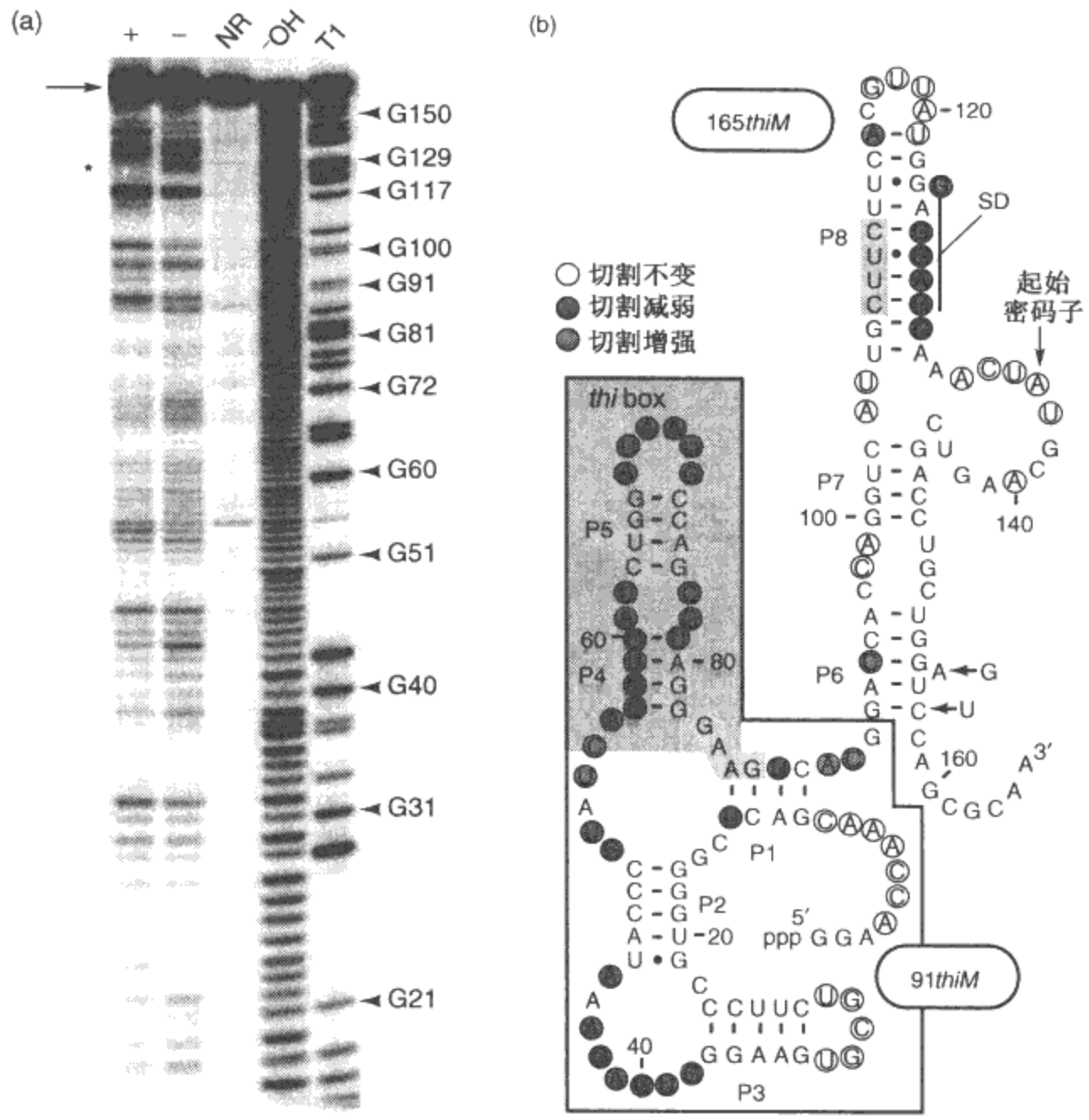


图 17.36　*thiM* mRNA 对 TPP 的结合。（a）165*thiM* mRNA 的串联探测。Breaker 及同事在 TPP 有（＋）或无（－）条件下，在 25℃孵育 165*thiM* mRNA 40h。NR 泳道的 RNA 没有孵育，$^{-}$OH 和 T1 表示 RNA 分别与碱基和 RNase 孵育。（b）预计的有 TPP 时 165*thiM* mRNA 的二级结构。Thi 盒用蓝色加亮，红色碱基表示经过 TPP 还原切割，黄色未配对碱基在切割时未变化。此处显示的橙色碱基 CUUC 与 Shine-Dalgarno 序列（SD）的 GGAG 配对，AGGA 是 CUUC 的另一个搭档。（*Source*：*Nature*，419，Wade Winkler，Ali Nahvi，Ronald R. Breaker，“Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression，” fig. 1 a&b，p. 953，Copyright 2002，reprinted by permission from Macmillan Publishers Ltd.）

mRNA 的二级结构也可用于控制翻译起始。*E. coli* 的 *thiM* mRNA 的 5′-UTR 含有一个核开关，包括一个核酸适体，它可结合硫胺及其代谢物硫胺磷酸，尤其是硫胺素焦磷酸（TPP）。当 TPP 丰富时，TPP 就与核酸适体结合引起 mRNA 构象改变，将 Shine-Dalgarno 序列隐藏在二级结构中。这种转换使 SD 序列不为核糖体所发现，从而抑制 mRNA 的翻译。这样可节省能量，因为 *thiM* mRNA 编码产生更多硫胺的酶，从而产生更多的 TPP。

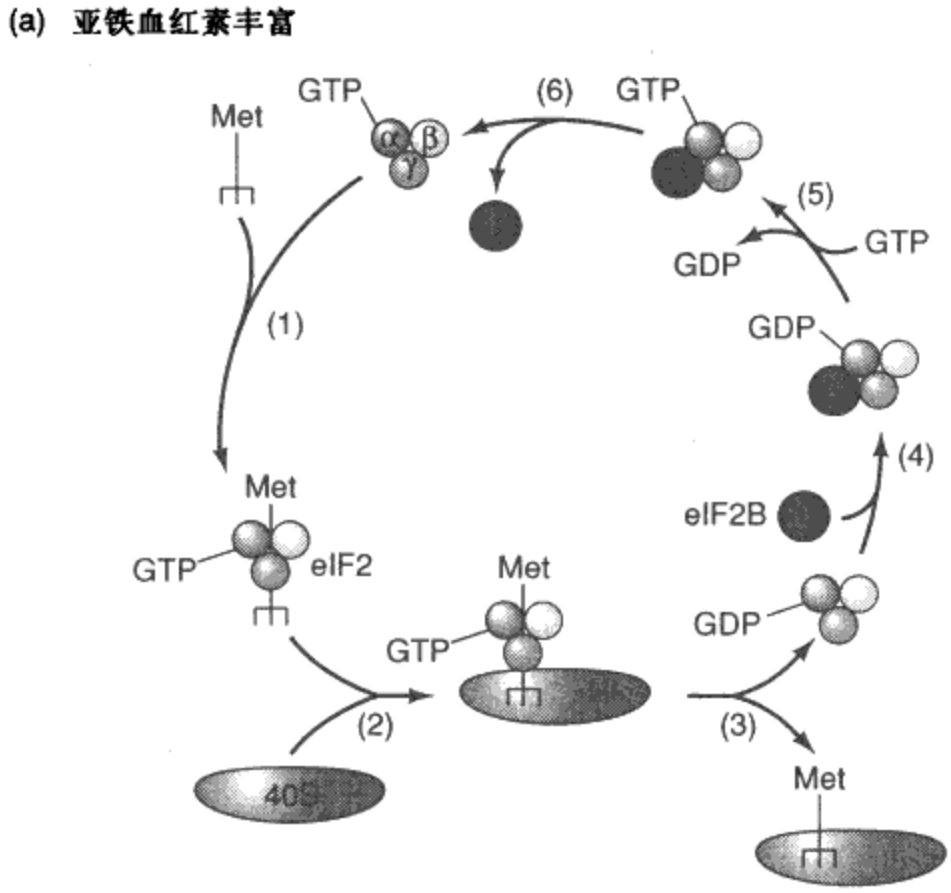
真核生物的翻译调控

真核生物 mRNA 的寿命比原核生物长，所以有更多的翻译调控机会。翻译中的限速因素通常是起始，所以我们期望找出在这一水平的最有效调控。这种调控的最普遍机制是起始因子的磷酸化，并且我们知道有些磷酸化在某个地方是抑制性的，而在另一个地方又是激活性的。最后，有一个例子是蛋白质直接结合到 mRNA 的 5′-UTR 阻止翻译，而去除这个蛋白质就激活翻译。

起始因子 eIF2α 的磷酸化　认识最清楚的抑制性磷酸化例子发生在网织红细胞（reticulocyte）中，该细胞只产生血红蛋白。但有时候网织红细胞缺乏血红素（血红蛋白的含铁部分），此时再继续生产血红蛋白的 α-β-球蛋白就是浪费了，但网织红细胞不

是停止产生球蛋白 mRNA，而是按以下方式阻断对它们的翻译（图 17.37）。血红素缺乏激发了一个被称为**血红素调控抑制物**（heme-controlled repressor，**HCR**）的蛋白激酶的活性。该酶使 eIF2 的一个亚基 **eIF2α 磷酸化**。eIF2 的磷酸化形式对 eIF2B 的结合比通常更紧密。eIF2B 是一个起始因子，其功能是使 eIF2 上的 GTP 交换为 GDP。若 eIF2B 对磷酸化的 eIF2 结合得太紧，就不能在其他 eIF2 分子上自由进行 GTP 和 GDP 的交换，这样 eIF2 就保持无活性的 GDP 结合形式，不能将 Met-tRNA$_i^{Met}$ 结合到 40S 核糖体上。因此，翻译起始立即停止。

抗病毒蛋白**干扰素**（interferon）遵循同样的途径。在干扰素和双链 RNA（出现在许多病毒感染里，但不在正常细胞里）存在时，另一个 eIF2a 激酶被活化。这个激酶被称作 **DAI**（double-stranded RNA-activated inhibitor of protein synthesis），表示**双链 RNA 激活的蛋白合成抑制物**。DAI 的作用与 HCR 相同，都是阻断翻译起始。这对病毒感染的细胞有益，因为病毒已经控制了细胞，阻断翻译将阻止子代病毒的繁殖，因而降低感染。



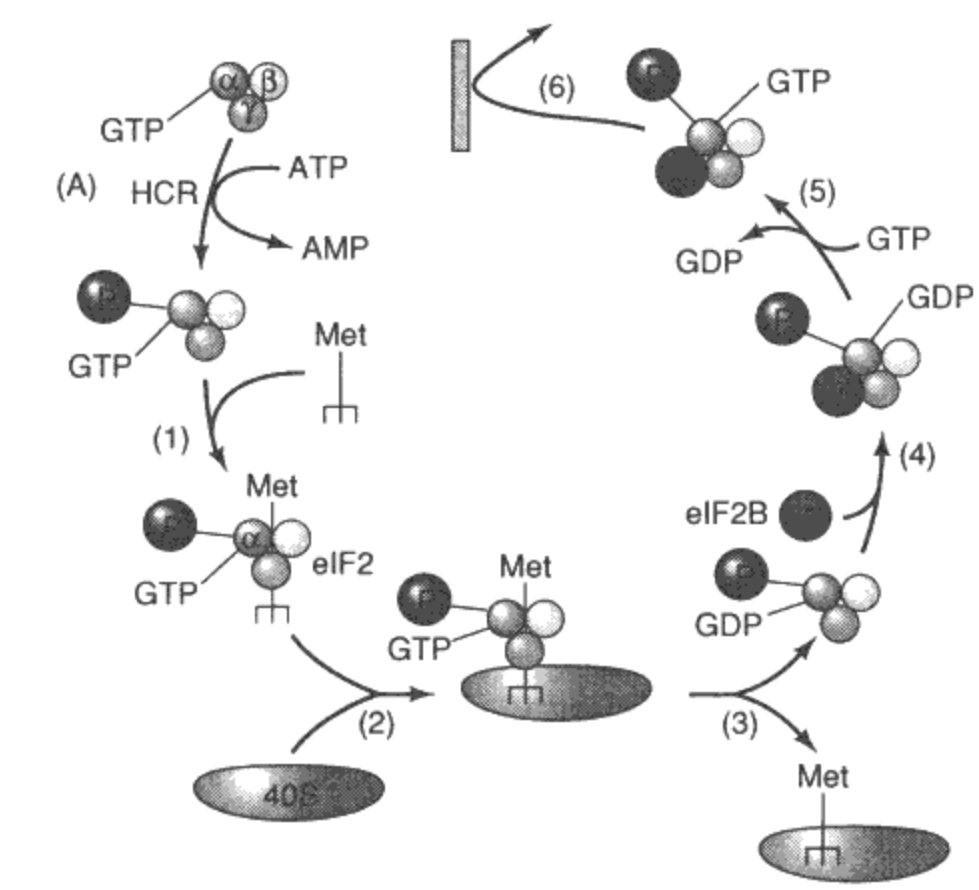


图 17.37　eIF2α 的磷酸化对翻译的抑制。（a）亚铁血红素丰富，没有抑制。步骤 1，Met-tRNA$_i^{Met}$结合到 eIF2-GTP 复合体，形成三元 Met-tRNA$_i^{Met}$ GTP-eIF2 复合物。eIF2 因子是一个不恒等亚基的 3 分子缩合物（α、β 和 γ）。步骤 2，三元复合物结合到 40S 核糖体亚基。步骤 3，GTP 水解成 GDP 和磷酸盐，使 GDP-eIF2 复合体从 40S 核糖体上游离下来，脱离 Met-tRNA$_i^{Met}$ 的黏附。步骤 4，eIF2B 结合到 eIF2-GDP 复合体。步骤 5，eIF2B 在复合物上用 GTP 交换 GDP。步骤 6，eIF2B 从复合物上游离下来。eIF2-GTP 和 Met-tRNA$_i^{Met}$能够聚集在一起形成新的复合物，开始新一轮的起始。（b）亚铁血红素缺乏导致翻译的抑制。步骤 A，HCR（被亚铁血红素缺乏所激活）连接一个磷酸基到 eIF2 的 α 亚基。步骤 1～5 与（a）的相同。但步骤 6 被阻止，因为 eIF2B 对磷酸化 eIF2α 的高亲和力，阻止其解离。eIF2B 被紧密地束缚在复合物上，翻译起始被抑制。

小结 真核生物 mRNA 的寿命相对较长，所以对翻译进行调控的机会比原核生物要多。eIF2 的 α 亚基是翻译调控的偏爱靶位。在血红素缺乏的网织红细胞中，HCR 是激活的，因此它能使 eIF2α 磷酸化，抑制翻译起始。在病毒感染的细胞中，另一种激酶 DAI 是活化的，它也使 eIF2α 磷酸化，抑制翻译起始。

eIF4E 结合蛋白的磷酸化 翻译起始中的限速步骤是帽子结合因子 eIF4E 对帽子的结合。有趣的是，eIF4E 也要被磷酸化，但其磷酸化不是抑制而是促进翻译的起始。磷酸化的 eIF4E 对帽子结合的亲和力是未磷酸化 eIF4E 的 4 倍，由此可解释其对翻译的促进作用。我们知道，有利于 eIF2α 磷酸化和翻译抑制的条件不利于细胞生长（如血红素缺乏和病毒感染），提示有利于 eIF4E 磷酸化和促进翻译的条件应该有利于细胞生长，一般来说这是正确的。用胰岛素或促细胞分裂素刺激细胞分裂可导致 eIF4E 磷酸化的增强。

胰岛素和多种生长因子，如血小板衍生生长因子（PDGF）通过另一个途径促进哺乳动物的翻译起始，该途径涉及 eIF4E。多年来我们已经知道，胰岛素和许多生长因子与细胞表面的专一性受体相互作用（图 17.38）。这些受体带有具酪氨酸激酶活性的胞内结构域，当与其配体相互作用时，这些受体能够二聚化，并自动磷酸化。换句话说，一个单体的酪氨酸激酶结构域，使另一个单体的酪氨酸磷酸化。由此引发了第 12 章所述的几个信号转导途径。其中一条途径激活 mTOR 蛋白，mTOR 的其中一个靶位是蛋白 PHAS-I。PHAS-I 结合 eIF4E，抑制其活性，具体讲是 PHAS-I 抑制 eIF4E 与 eIF4G 的结合。一旦被 mTOR 磷酸化，PHAS-I 就从 eIF4E 上游离下来，再自由地与 eIF4G 结合，促进 mRNA 和 40S 核糖体亚基活性复合物的形成（图 17.38 和图 17.27），因此增强翻译。

1994 年，Sonenberg 及同事用 Far Western 筛选结合 eIF4E 的蛋白质时，发现了人的 PHAS-I。Far Western 筛选方法除了探针用普通的标记蛋白质代替抗体以外，与抗体表达库的筛选（第 4 章）相似。所以，这里是寻找两个非抗体蛋白质的相互作用，而不是一个蛋白质被一个抗体的识别。本例中，研究者用 eIF4E 衍生物探测了人表达文库（在 λgt11 里），寻找 eIF4E 结合蛋白。探针是 eIF4E，耦合了心肌激酶（heart muscle kinase，HMK）的磷酸化位点，该位点随后被 [γ-^{32}P] 标记。在筛选的上百万个噬菌斑中，有 9 个含有编码结合 eIF4E 探针的蛋白质基因，其中 3 个含有编码 eIF4F 的 eIF4G 亚基的至少部分基因，因此，它们结合 eIF4E 并不奇怪。另外 6 个阳性克隆编码两个相关蛋白质：4E-BP1 和 4E-BP2（eIF4E-binding protein 1 and 4E-BP 2，表示 eIF4E 结合蛋白 1 和 2）。4E-BP1 蛋白与先前发现的大鼠 PHAS-I 蛋白同源。

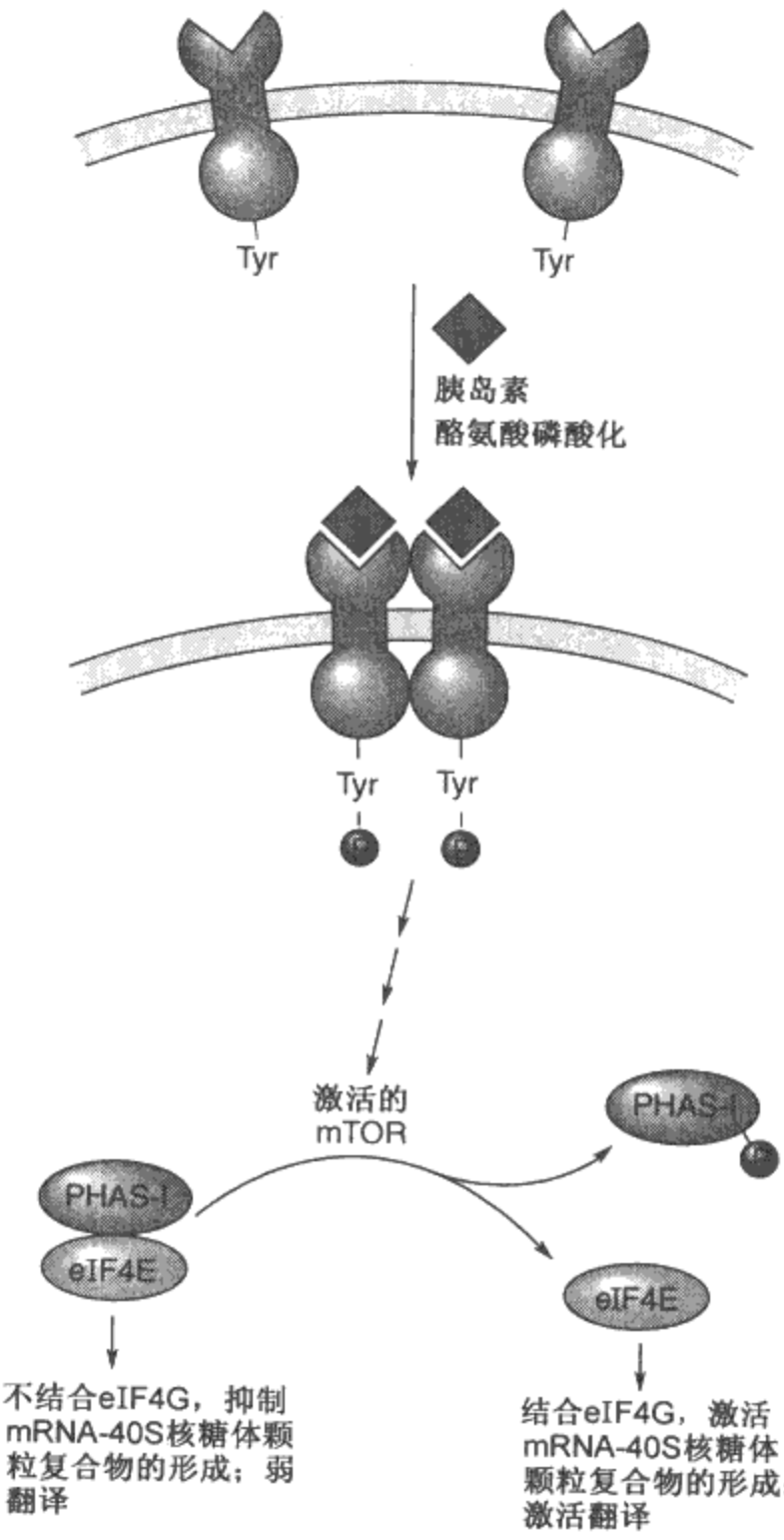


图 17.38 PHAS-I 磷酸化对翻译的激活。胰岛素或生长因子（如 EGF）与其在细胞表面的受体结合。通过一系列步骤活化了蛋白激酶 mTOR。mTOR 的其中一个靶蛋白是 PHAS-I。PHAS-I 被 mTOR 磷酸化后与 eIF4E 解离，释放 eIF4E 使其与 eIF4G 结合，参与激活翻译起始。

为了证明 4E-BP1、4E-BP2 和 PHAS-I 确实结合 eIF4E，研究人员表达了 4E-BP1H 和 4E-BP2 与 GST（谷胱甘肽-S-转移酶）的融合蛋白，PHAS-I 与寡聚组氨酸（第 4 章）的融合蛋白。然后做 Far Western 印迹，将融合蛋白进行 SDS-PAGE，再印迹到硝酸纤维素膜上，像以前一样用放射性标记的

HMK-eIF4E 做探针。图 17.39 显示 GST 单独不能结合 eIF4E 探针，但是所有其他蛋白质都能与之结合，因此这 3 种蛋白质确实结合了 eIF4E。

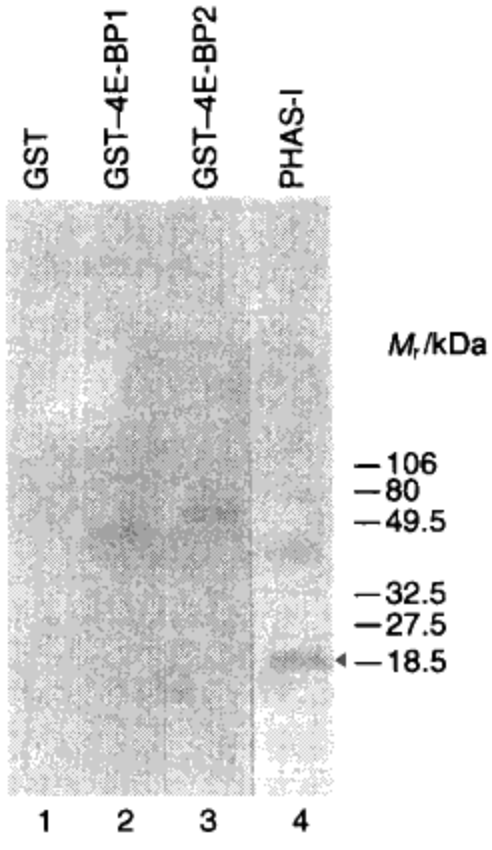


图 17.39　eIF4E 与 4E-BP1、4E-BP2、PHAS-I 之间的结合。GST、GST-4E-BP1、GST-4E-BP2 和寡聚组氨酸-PHAS-I 在上方标出，对它们进行 SDS-PAGE，然后印迹到硝酸纤维素膜上，用 ^{32}P 标记的 HMK-eIF4E 探测。GST 未与探针结合，但其他蛋白质都结合了，证明它们都结合到 eIF4E 上了。（*Source*：Pause，A.，G. J. Belsham，A.-C. Gingras，O. Donzé，T.-A. Lin，J. C. Lawrence，and N. Sonenberg，Insulin-dependent stimulation of protein synthesis by phosphorylation of a regulator of 5′-cap function. *Nature* 371（27 Oct 1994）p. 763，f. 2. Copyright © Macmillan Magazines Ltd.）

因为已知 PHAS-I 磷酸化是对胰岛素做出的响应，那么了解这种磷酸化是否改变 4E-BP1 和 eIF4E 之间的相互作用就很重要。为了找出答案，研究者分别提取了胰岛素处理和未处理的脂肪组织，然后在 m^{7}GDP-琼脂糖上进行亲和层析，之后将蛋白质印迹到硝酸纤维素膜上，用同时抗 eIF4E 和 4E-BP1 的抗体探测该 Western 膜。图 17.40 显示，胰岛素处理细胞的 eIF4E 对帽子同系物树脂的结合与未处理细胞 eIF4E 的一样紧密。但是胰岛素处理细胞的 4E-BP1 对帽子同系物树脂的结合却极大地降低了，因为 4E-BP1 对树脂的结合是通过其与 eIF4E 的相互作用实现的。这一结果明显地提示胰岛素处理细胞中，4E-BP1 和 eIF4E 之间的结合变弱了。

接下来，研究者证明了 4E-BP1 和 4E-BP2 均能

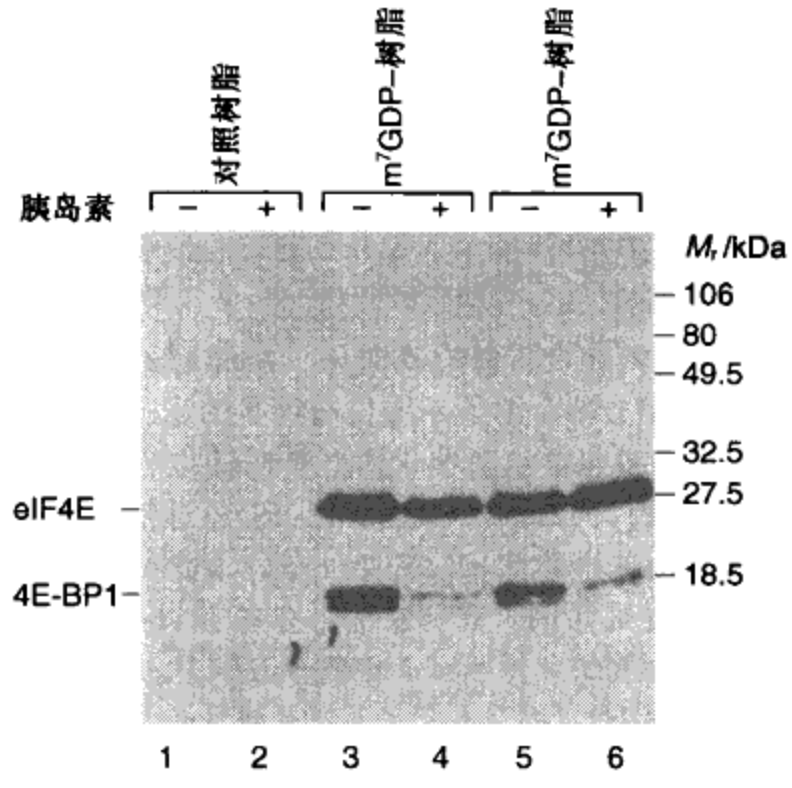


图 17.40　胰岛素对 eIF4E 与 4EBP1 之间结合的影响。Lawrence 及其合作者将大鼠脂肪组织分别在有或无胰岛素条件下孵育，然后进行组织匀浆，通过结合到 m^{7}GDP 琼脂糖上纯化提取物中的帽子结合蛋白。然后用 SDS 除去树脂中的帽子结合蛋白，SDS-PAGE 凝胶电泳将其除去，然后印迹到硝酸纤维素膜上，用直接对 eIF4E 和 4E-BP1 的抗体进行探测，用标记的二抗检测抗体。泳道 1、2 表明不含 m^{7}GDP 的对照树脂上无结合物，泳道 3～6 中的蛋白质结合到 m^{7}GDP 琼脂糖上，这些蛋白质分别来自胰岛素处理（+）和未处理（–）组织，如图上部标示。eIF4E 和 4E-BP1 的位置在左边标出。胰岛素处理极大地减少了与帽子同系物树脂结合的 4E-BP1 的量，这一结果证明胰岛素处理削弱了 eIF4E 和 4E-BP1 之间的相互作用。（*Source*：Pause，A.，G. J. Belsham，A. C. Gingras，O. Donzé，J.-A. Lin，J. C. Lawrence，and N. Sonenberg，Insulin-dependent stimulation of protein synthesis by phosphorylation of a regulator of 5′-cap function. *Nature* 371（27 Oct 1994）p. 765，f. 4c. Copyright © Macmillan Magazines Ltd.）

抑制体内和体外的帽子依赖性翻译途径，但不能抑制帽子非依赖性翻译途径。他们用了一个加帽的含有两个报告基因转录物的双顺反子 mRNA（图 17.41a），其 5′ 端有一个氯霉素乙酰基转移酶（CAT），CAT 的翻译是帽子依赖性的。另一个是萤光素酶，其翻译由细小核糖核酸病毒的内部核糖体进入位点（IRES）驱动。CAT 和萤光素酶都是很好的报告基因（第 5 章），它们的产物可以方便灵敏地检测到。实验中以萤光素酶的合成作为内标，因为

不管帽子依赖性翻译是否进行，其合成都会发生。图 17.41b 显示体外检测兔网织红细胞提取物中 CAT 和萤光素酶的翻译结果。Sonnenberg 和 Lawrence 在不同时间把 4E-BP1 和 4E-BP2（GST 融合蛋白）加入到无细胞提取物中。正如所预期的，萤光素酶产物自始至终相对恒定，但加入 GST-4E-BP1 或 GST-4E-BP2 中任意一个时，CAT 产物显著减少。如果在加入 GST-4E-BP1 或 GST-4E-BP2 的同时或提前 10min 加入 eIF4E，则不会对 CAT 翻译产生抑制，由此提示多余的 eIF4E 足以吸收 GST-4E-BP1 或-4E-BP2，阻止它对翻译起始的抑制。在体内他们也得到了类似的结果，将带有两种报告基因的质粒和编码 4E-BPI 和 4E-BP2 的质粒一起转入细胞，检测两种报告基因的产物。在有 4E-BP1 或 4E-BP2 的情况下，CAT 的合成被抑制约 5 倍，而萤光素酶的合成却未受影响。

在另一个实验中，Lawrence 和 Sonnenberg 及同事直接测定了 PHAS-I 磷酸化的效应。用 MAP 激酶在体外使 PHAS-I 磷酸化（实际上是［His_6］PHAS-I），然后检测它与 eIF4E 结合的能力。图 17.42 显示，未被磷酸化的 PHAS-I 与固定到m^7GTP-琼脂糖凝胶上的 eIF4E 结合（泳道 7），而磷酸化的 PHAS-I 则不能与之结合（泳道 8）。最后，磷酸化的 PHAS-I 又通过磷酸酯酶 PP2A 去磷酸化之后，能像未磷酸化的 PHAS-I 那样与 eIF4E 结合（泳道 9）。所以，PHAS-I 的磷酸化减弱了 PHAS-I 和 eIF4E 间的相互作用。实验中采用的是 MAP 激酶，虽然现在认为 mTOR 才是 PHAS-I 的体内磷酸化酶。

小结　胰岛素和一些生长因子激活蛋白激酶 mTOR 参与的途径。mTOR 激酶的靶蛋白之一是 PHAS-I（大鼠）或 4E-BP1（人）。在 mTOR 的磷酸化作用下，该蛋白质从 eIF4E 上解离下来，释放 eIF4E 参与更活跃的翻译起始。

通过 eIF4E 结合蛋白 Maskin 调控翻译的起始

真核细胞也可利用其他蛋白质来影响 eIF4E 从而抑制翻译起始，其中一种蛋白质是在非洲爪蟾（*Xenopus laevis*）中发现的，叫做 **Maskin**。图 17.43 是目前对 Maskin 如何抑制爪蟾卵母细胞中细胞周期蛋

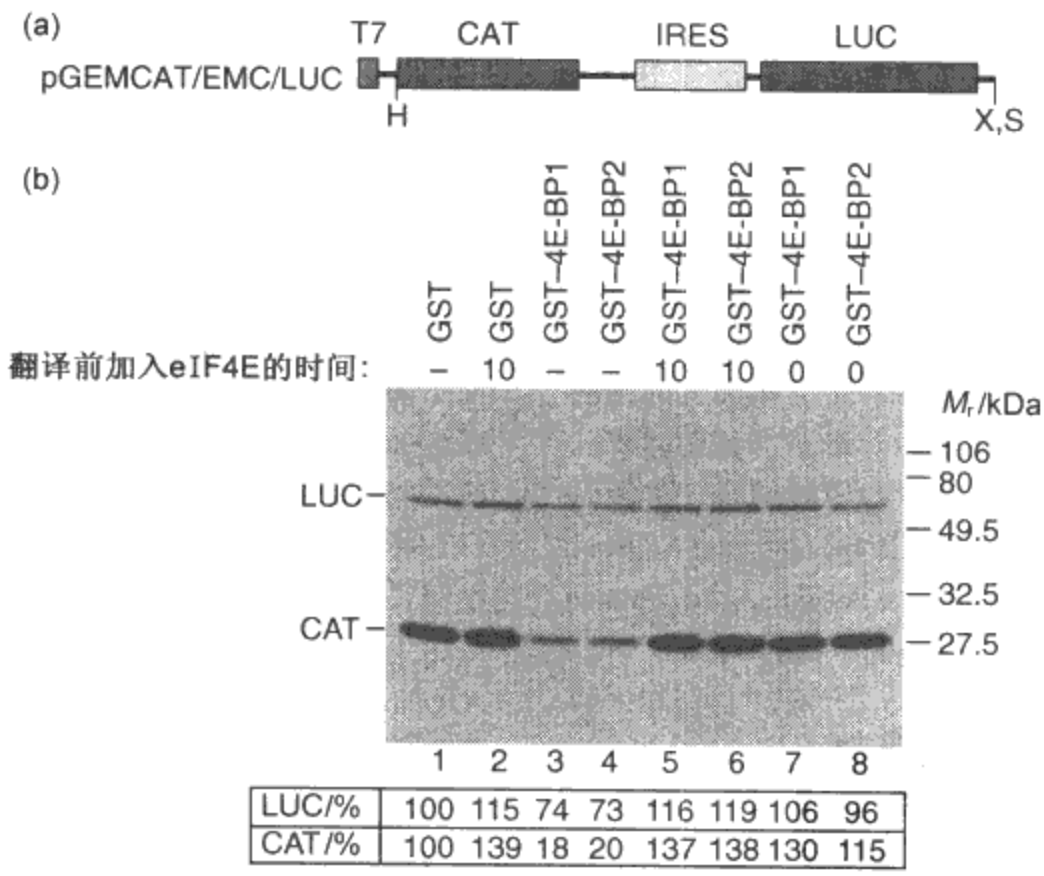


图 17.41　在体外 4E-BP1 和 4E-BP2 对帽子依赖性和帽子非依赖性翻译的影响。（a）双顺反子的构建。Lawrence 及同事通过 T7 RNA 聚合酶在体外对质粒（图中所示）进行转录构建了双顺反子 RNA，在其 5′端加上帽子。CAT 顺反子的翻译是帽子依赖性的，而萤光素酶顺反子（LUC）的翻译由 IRES 驱动。（b）实验结果。将加帽 RNA 与其他成分（图上部标注）一同加入到兔网织红细胞提取物中。泳道 3 和 4 表明 4E-BP 融合蛋白抑制帽子依赖性 CAT 的翻译，但不抑制帽子非依赖性 LUC 的翻译。泳道 5～8 表明不管在融合蛋白之前 10min 还是同时加入，eIF4E 都可逆转这种抑制作用。（*Source*: Pause, A., G. J. Belsham, A.-C. Gingras, O. Donzé, T.-A. Lin, J. C. Lawrence, and N. Sonenberg, Insulin-dependent stimulation of protein synthesis by phosphorylation of a regulator of 5′-cap function. *Nature* 371 (27 Oct 1994) p. 766, f. 5a (bottom). Copyright © Macmillan Magazines Ltd.)

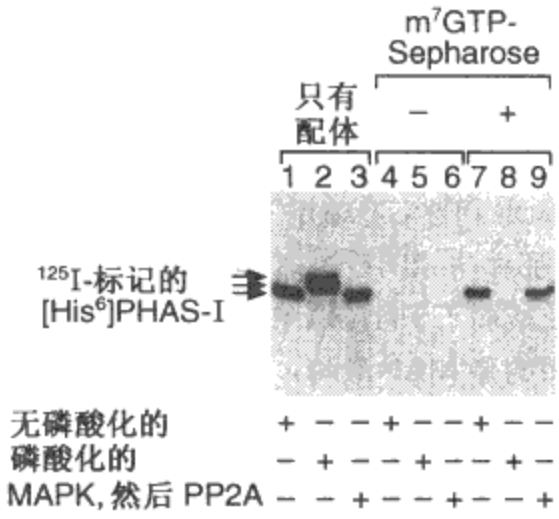


图 17.42　PHAS-I 磷酸化对结合 eIF4E 的作用。Lawrence 和 Sonnenberg 及同事制备了重组 His-标签的大鼠 PHAS-I [$(His)_6$ PHAS-I]，然后以 3 种不同方式处理：①无修饰；②在体外用 MAP 激酶磷酸化；③先用 PP2A 蛋白激酶磷酸化接着再去磷酸化。然后用^{125}I 标记这 3 种处理的 PHAS-I，检测它们分别对不同处理 m^7GTP-琼脂糖凝胶树脂的结合能力，凝胶与胰岛素处理细胞提取物共孵育（+）和不孵育（–）。对从树脂上洗脱的蛋白质进行 SDS-PAGE。泳道 1～3（只有配体）显示过柱前 3 种形式的 [$(His)_6$]-PHAS-I；泳道 4～6 显示 3 种形式的 [$(His)_6$]-PHAS-I 都不能与无细胞提取物的柱子结合，因此这 3 种 PHAS-I 样品自身都没有 m^7GTP 树脂亲和性；泳道 7 和 9 显示，两种未磷酸化的 PHAS-I 可与胰岛素处理细胞提取物所孵育的柱子结合，这种结合的基础可能是 PHAS-I 结合到提取物的 eIF4E 上，而后者又与树脂上的帽同系物（m^7GTP）结合了。此外，泳道 8 证明体外已经磷酸化的 PHAS-I 样品不能与树脂结合，推测可能是不能与树脂上固定的 eIF4E 结合。（*Source*: Lin, T.-A., PHAS-I as a link between mitogen-activated protein kinase and translation initiation. *Science* 266 (28 Oct 1994) p. 655, f. 3b. Copyright © AAAS.）

白 B 的 mRNA 翻译所做的假设。我们在第 15 章已经知道，爪蟾卵母细胞许多 mRNA 的 poly（A）尾非常短，不能被很好地翻译。原因之一可能是细胞质多聚腺苷酸元件（cytoplasmic polyadenylation element，CPE）被一个结合蛋白 **CPEB** 占据，该蛋白质又与 Maskin 结合，而 Maskin 又结合到 eIF4E 上。在这种相互作用中，Maskin 与 PHAS-I 一样，阻断了 eIF4E 和 eIF4G 之间的相互作用，从而抑制翻译起始。

当 *Xenopus* 卵母细胞被激活后，CPEB 被 Eg2 酶磷酸化，这种磷酸化表现出两种主要作用。首先，它吸引切割与多聚腺苷酸化特异因子 CPSF（cleavage and polyadenylation specificity factor）识别 mRNA的多聚腺苷酸信号（AAUAAA），由此激活休眠态 mRNA 的多聚腺苷酸化。其次，CPEB 的磷酸化（或由该磷酸化引起的多聚腺苷酸化）明显地使 CPEB 与 eIF4E 解离，而使 eIF4E 与 eIF4G 结合，激发翻译的起始。

细胞周期蛋白 B（Maskin 所调控基因中的一个）是细胞周期的关键激活因子。因此像细胞分裂这样的基本过程要受翻译水平上的调控。

小结　在 *Xenopus* 卵母细胞中，Maskin 结合到 eIF4E 和结合有休眠细胞周期蛋白 B mRNA 的 CPEB 上。由于 Maskin 的结合，eIF4E 不能与 eIF4G 结合，所以翻译受到抑制。在卵母细胞活化时，CPEB 被磷酸化，由此激活多聚腺苷酸化，有可能引起 CPEB 从 Maskin 上解离，Maskin 随后又从 eIF4E 上解离。因为 Maskin 不再附着，eIF4E 就可以自由地与 eIF4G 结合，翻译起始。

mRNA 结合蛋白的促进作用　我们已经知道 mRNA 二级结构可以影响细菌基因的翻译，这在真核生物也是如此。让我们来看一个研究得很清楚的由 RNA 二级结构元件（茎环）与 RNA 结合蛋白之间相互作用对 mRNA 翻译抑制的例子。在第 16 章我们已学过，转铁蛋白（transferrin）受体和铁蛋白（ferritin）两种铁相关蛋白的浓度受铁离子浓度调控。当血清中铁离子浓度高时，由于编码转铁蛋白受体的 mRNA 不稳定，导致转铁蛋白受体的合成速度下降。同时，作为胞内铁储存蛋白的铁蛋白的合成增加。铁蛋白由 L 和 H 两条多肽链组成。铁离子使编码铁蛋白两个肽链 mRNA 的翻译水平增加。

什么原因引起了翻译效率的增加呢？两个研究小组几乎同时对这个问题得出了一致的结论。第一组由 Hamish Munro 领导，检测了大鼠铁蛋白 mRNA 的翻译情况；第二组由 Richard Klausner 领导，研究了人铁蛋白 mRNA 的翻译。回想第 16 章转铁蛋白受体 mRNA 的 3′-UTR 含有若干个茎环结构，称为铁离子应答元件（IRE），能够与蛋白质结合。我们也知道铁蛋白 mRNA 的 5′-UTR 有一个非常类似的 IRE。而且在脊椎动物中，铁蛋白的 IRE 高度保守，甚至比基因本身的编码区还要保守。这些观察结果有力地提示，铁蛋白的 IRE 在铁蛋白 mRNA 的翻译中发挥作用。

为了验证该预测，Munro 及同事构建了含有 CAT 报告基因的 DNA 质粒，其两侧是大鼠铁蛋白 L 基因的 5′-UTR 和 3′-UTR。在质粒（pLJ5CAT3）中，CAT 的转录由非常强的反转录病毒的启动子-增强子驱动。在另一个质粒（pWE5CAT3）中，CAT

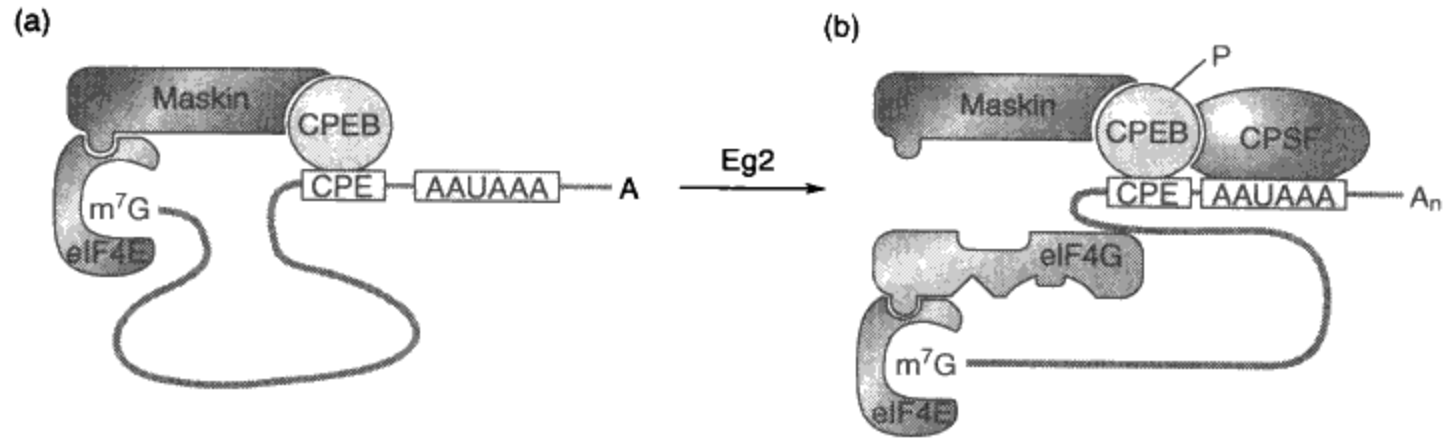


图 17.43　Maskin 调控的翻译起始模型。（a）在休眠的爪蟾卵母细胞中，CPEB 与细胞周期蛋白 B mRNA 上的 CPE 结合，Maskin 被结合到 CPEB 上，eIF4E 被结合到 Maskin 上。后一步相互作用干扰 eIF4E 与 eIF4G 结合的能力，而 eIF4E 与 eIF4G 的结合是翻译起始所必需的。结果，细胞周期 B 蛋白 mRNA 处于休眠状态。（b）在活化状态，Eg2 使 CPEB 磷酸化，允许 CPSF 结合，mRNA 多聚腺苷酸化，这也明显地引起 Maskin 从 eIF4E 上解离，从而使 eIF4E 与 eIF4G 结合，激发翻译的起始。（*Source*：Adapted from Richter，J. D. and W. E. Theurkauf，The message is in the translation. *Science* 293［2001］p. 61，f. 1.）

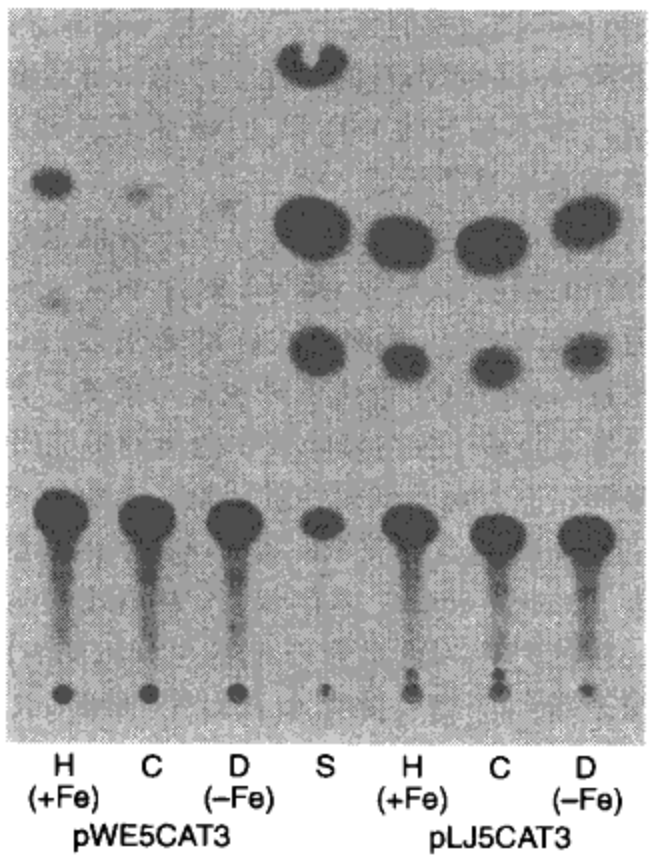


图 17.44　铁离子对重组 5CAT3 翻译阻遏的解除。Munro 及同事制备了两个重组质粒，其中 CAT 报告基因的两侧是大鼠铁蛋白 L 基因的 5′-UTR 和 3′-UTR。把这两种质粒转入细胞，分别置于弱启动子（在 pWE5CAT3 是 β-肌动蛋白启动子）或强启动子（在 pLJ5CAT3 是反转录病毒启动子-增强子）的调控下。泳道 H 的细胞用血晶素处理，泳道 D 的细胞用铁离子螯合剂-去铁胺甲磺酸盐（desferal）去除铁离子。泳道 C 的细胞未经处理。检测每组细胞的 CAT 活性（如第 5 章所述）。泳道 S 是标准 CAT 反应，显示出氯霉素底物及其乙酰化形式的位置。左边的泳道显示当 CAT mRNA 不足时，其翻译可通过铁离子诱导。作为对照，右边的泳道显示当 mRNA 丰度很高时，其翻译不受铁离子诱导。（*Source*：Adapted from Aziz，N. and H. N. Munro，iron regulates ferritin mRNA translation through a segment of its 5′untranslated region. *Proceedings of the National Academy of Sciences USA* 84（1997）p. 8481，f. 6.）

的转录受弱的 β-肌动蛋白启动子调控。接着，他们把这些 DNA 转入哺乳动物细胞，并在有铁源［血晶素（hemin）］、铁离子螯合剂［甲磺酸去铁胺（desferal）］或无添加物的情况下，对 CAT 产物进行检测。图 17.44 结果显示，当细胞携带 WE5CAT3 质粒中的 CAT 基因时，CAT mRNA 的量相对很少。在这种情况下，CAT 产量低，但可受铁离子（比较左边泳道的 C 和 H）诱导和铁离子螯合剂（比较左边泳道的 C 和 D）的抑制。相反，当细胞携带 pLJ5CAT3，CAT mRNA 的量相对丰富时，CAT 产量高且不可诱导。对该结果最简单的解释是阻遏物结合在铁蛋白 5′-UTR 的 IRE 上，阻断了与之相连的 CAT 顺反子的翻译。铁离子能以某种机制去除阻遏物，允许翻译发生。当 CAT mRNA 丰富时，CAT 的合成不被诱导，因为 mRNA 分子的数量极大地超过了阻遏物分子。几乎没有抑制时，观察不到诱导。

如何知道 IRE 参与了阻遏呢？我们怎样知道是 5′-UTR 而非 3′-UTR 更重要呢？Munro 及同事回答了这个问题，他们制备了两个新质粒，一个含 5′-UTR 但缺失 3′-UTR，另一个包含两个 UTR，但缺失包含 IRE 的前 67nt。图 17.45 显示缺失铁蛋白 mRNA 3′-UTR 的 pWE5CAT 仍然维持 CAT 的铁离子诱导活性。而缺少 IRE 的 pWE5sCAT3 在加入或不加铁离子时，CAT 的表达量都很高。这一结果不仅说明 IRE 负责诱导，而且进一步证实了 IRE 介导阻遏，因为 IRE 的缺失导致 CAT 甚至在没有铁离子存在时也能高水平表达。

可以做出结论，一些阻遏物蛋白必须结合到铁蛋白 mRNA 5′-UTR 的 IRE 上产生阻遏作用，直到铁离子以某种机制将其去除。由于铁蛋白 mRNA、铁传递蛋白 mRNA 上的 IRE 高度保守，预测至少其中有一些蛋白质可能在两种情况下都起作用。事实上，正如在第 16 章学过的，顺乌头酸酶脱辅基蛋白

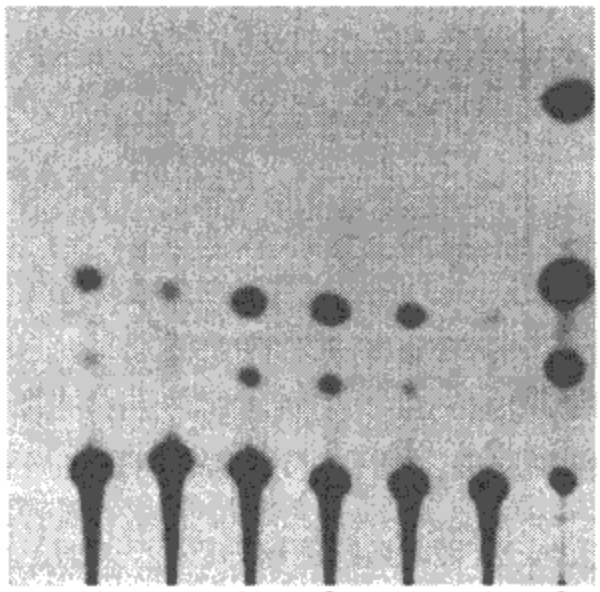


图 17.45　pWE5CAT3 的 5′-UTR 中 IRE 对于铁离子诱导性的重要性。Munro 及同事用母本质粒，如图 17.44 所示的 pWE5CAT3 及其两个衍生物 pWE5sCAT3、pWE5CAT 转染细胞。pWE5sCAT3 缺失铁蛋白 5′-UTR 前 67nt 包括 IRE，pWE5CAT 缺失铁蛋白 3′-UTR。这些细胞用血晶素处理或不处理，然后分析每批细胞的 CAT 活性。IRE 的缺失会引起铁离子诱导性的丧失。（*Source*：Adapted from Aziz，N. and H. N. Munro，Iron regulates ferritin mRNA translation through a segment of its 5′-untranslated region. *Proceedings of the National Academy of Sciences USA* 84（1987）p. 8482，f. 7.）

（aconitase apoprotein）是 IRE 结合蛋白，当它与铁离子结合后，从 IRE 上解离下来，解除抑制。

小结　铁蛋白 mRNA 的翻译受铁离子诱导。这种诱导作用按以下方式进行：阻遏蛋白（顺乌头酸酶脱辅基蛋白）结合到铁蛋白 mRNA 5′-UTR 区 5′端附近的茎环铁离子应答元件（IRE）上。铁离子使阻遏物离开，允许 mRNA 翻译继续进行。

miRNA 对翻译起始的阻断　在第 16 章我们已经了解到 miRNA 以两种方式控制基因的表达：当与其靶 mRNA 完美配对时引起 mRNA 降解；如果碱基配对不完全，可通过一种尚未解释的机制抑制蛋白质的产生。Witold Filipowicz 及同事希望揭开这个神秘的机制，并于 2005 年发表了研究结果，表明不完全配对的哺乳动物 *let-7* miRNA 可能是通过干扰帽子的识别而抑制翻译的起始。

他们用两种报告基因做探针，分别是海肾（*Renilla reniformis*）萤光素酶（RL）基因和萤火虫萤光素酶（FL）基因，它们的产物很容易分析：将荧光素与 ATP 混合时，可产生光。报告基因的 3′-UTR 构建成与 *let-7* miRNA 完全配对（Pref）或互补区有 1 个或 3 个不配对区引起 miRNA-mRNA 双链凸出的形式。这些改变的基因分别命名为 1×Bulge 和 3×Bulge，野生型对照基因（*Con*）对 *let-7* miRNA 无互补性。

当 Filipowicz 及同事用报告基因转化人细胞时，发现 RL-Pref 和 RL-3×Bulge 基因的表达与对照相比迅速降低（达 10 倍），而且这种降低可以被共转染的竞争 RNA（与 *let-7* miRNA 互补）所阻断，说明正如我们所预期的那样，miRNA 参与了基因表达的调控。

根据第 16 章的范例，我们预期 RL-Pref mRNA 的量会降低，因为 mRNA 与 miRNA 的完全配对会导致 mRNA 的降解。Filipowicz 及同事确实观察到该 mRNA 表达量降低为原先的 1/5 。而对于 RL-3×Bulge mRNA，我们预计其表达量不会显著降低，因为 mRNA 与 miRNA 的不完全配对会导致翻译受到干扰，而不是导致 mRNA 的破坏。事实上，该 mRNA 的表达只降低了 20%。

这些结果与 RL-3×Bulge 表达的降低被解释为由于翻译的阻断而不是 mRNA 的降解这一假说一致。但是 miRNA 以某种方式瞄准新生蛋白质使蛋白质水解，从而引起表达下降也是可能的。如果是这样，那么将新生蛋白质隐藏在内质网中（ER）可免于降解，不会观察到表达的降低。为验证这一假说，Filipowicz 及同事将 RL-3×Bulge 基因与红细胞凝集素（hemaglutinin）基因偶联，融合蛋白的 N 端表达一段信号序列，指导新生蛋白到 ER 的内腔。该重组子蛋白产物的降低量相对于对照与 RL-3×Bulge 自身的降低量是一样的。因此，*let-7* miRNA 的靶标是蛋白质的合成而不是蛋白质产物自身。

那么，*let-7* miRNA 阻止了翻译过程的哪一步呢？为了回答这个问题，Filipowicz 及同事收集了 RL-3×Bulge 基因转化细胞的多核糖体（polysome，由多个核糖体翻译的 mRNA，见第 18 章）。为了检测 RL-3×Bulge mRNA 在多核糖体中的情况，他们对多核糖体的各组分进行了 Northern 分析（图 17.46）。在给定 mRNA 上，翻译的起始越活跃，就会有越多的核糖体黏附在 mRNA 上，因此多核糖体就越重。在图 17.46 中最重的多核糖体在图的右边，可以清楚地看到对照 RL mRNA 存在于比 RL-3×Bulge mRNA（图 17.46b）大得多的核糖体中（图 17.46a 的更右边），将这些结果作图见图 17.46c。这种多核糖体分布的变动当与阻断 miRNA-mRNA 相互作用的反义-*let-7* miRNA 共转染时可被消除，在 RL-3×Bulge mRNA 与 miRNA 杂交的 3′-UTR 被突变移去时也可被消除。以上结果表明 RL-3×Bulge

mRNA 翻译的起始相对于对照 mRNA 的起始被显著地抑制了。因此，翻译起始（核糖体对mRNA的结合）好像是 *let-7* miRNA 影响翻译的靶位。

进一步的研究表明，mRNA 的 poly（A）尾在 *let-7* miRNA 对翻译的抑制中不起作用，poly（A）$^+$ 和 poly（A）$^-$ mRNA 的翻译同等程度地受 *let-7* miRNA 的抑制。但是帽子却起到很大作用，如我们所见，无帽 mRNA 的翻译非常弱。所以 Filipowicz 及同事给 RL 或 FL mRNA 加上了脑心肌炎病毒（encephalomyocarditis virus，EMCV）的内部核糖体进入位点（IRES），该病毒允许帽子不依赖性翻译。然后比较 *let-7* miRNA 对帽子依赖性和帽子不依赖性翻译的影响。与正常情况一样，*let-7* 抑制帽子依赖性 FL-3×Bulge mRNA 的翻译，而对带有 EMCV IRES 的帽子不依赖性 FL-3×Bulge mRNA 的翻译无影响。因此，*let-7* miRNA 看起来以帽子依赖性翻译的起始为靶向。

为了确定被 *let-7* miRNA 影响的帽子依赖性起始的具体部位，Filipowicz 及同事构建了双顺反子质粒，其中包含一个 eIF4E 或 eIF4G mRNA，它们被束缚于 RL 顺反子之前的顺反子间（intercistronic）区域。做法如下（图 17.47a）：在顺反子间区域，置入 2 个叫做 BoxB 的茎环，该茎环与 N 肽具有亲和性。然后，重组 eIF4E 和 eIF4G 基因分别插入了 N 肽-红细胞凝集素编码区，使这两个起始因子都被表达为带有 N 肽标签的融合蛋白。这些融合蛋白与 BoxB 茎环结合，从而激发双顺反子上 RL 顺反子的翻译。FL 顺反子的翻译是帽子依赖性的，因为该顺反子在加帽 mRNA 中首先出现。但是，只要其中一个起始因子被束缚于顺反子间区，RL 顺反子的翻译就是帽非依赖性的，因此该起始因子显然吸引了起始所需的所有其他因子。

接着，Filipowicz 及同事检测了该融合基因 FL 和 RL 部分的表达，用的是对照 3′-UTR，或 3×Bulge 的 3′-UTR，及束缚于顺反子间区域的起始因子 eIF4E 或 eIF4G（或作为阴性对照的 *lacZ* 产物 β-半乳糖苷酶）。图 17.47 显示实验结果。正如预期的，FL 顺反子的翻译是帽子依赖性的，与对照 mRNA

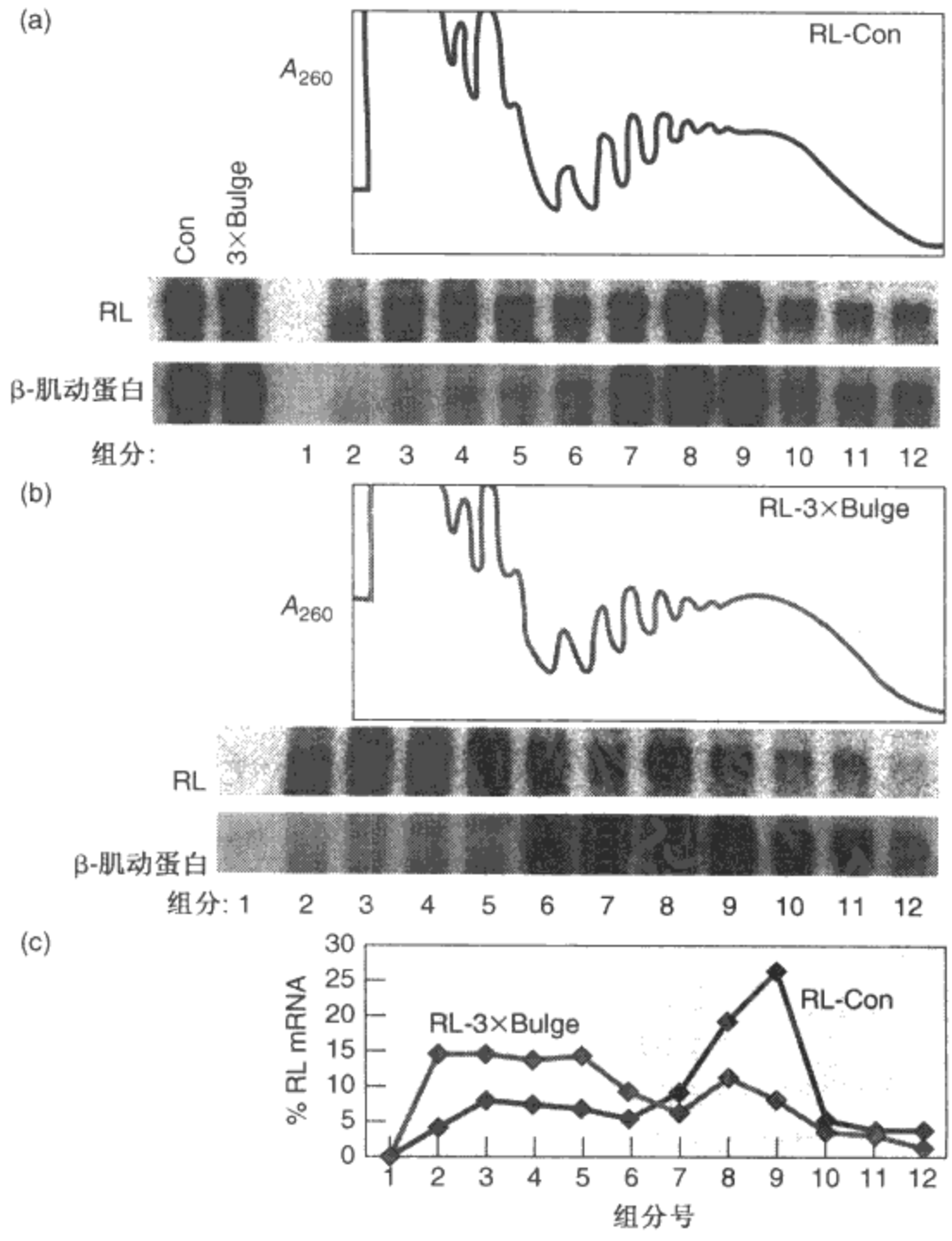


图 17.46　RL mRNA 的多核糖体的分布。Filipowicz 及同事用编码（a）对照 RL mRNA（RL-Con）或（b）RL-3×Bulge mRNA 的基因转染人类细胞。然后用蔗糖梯度离心展示多核糖体，将多核糖体分布中不同组分的 RNA 进行 Northern 印迹，印迹膜用放射性探针探测 RL 或 β-actin mRNA。后者是一个通常的细胞 mRNA，用作阳性对照。图（a）中 Northern 膜上最左边两个泳道的 RNA 来自超速离心前的样品。（c）显示对照和 RL-3×Bulge 多核糖体分布各组分的总放射性的百分比。（*Source*：（a-c）Reprinted with permission from *Science*, Vol. 309, Ramesh S. Pillai, Suvendra N. Bhattacharyya, Caroline G. Artus, Tabea Zoller, Nicolas Cougot, Eugenia Basyuk, Edouard Bertrand, and Witold Filipowicz, "Inhibition of Translational Initiation by Let-7 microRNA in Human Cells" Fig. 1 c&e, p. 1574, Copyright 2004, AAAS)

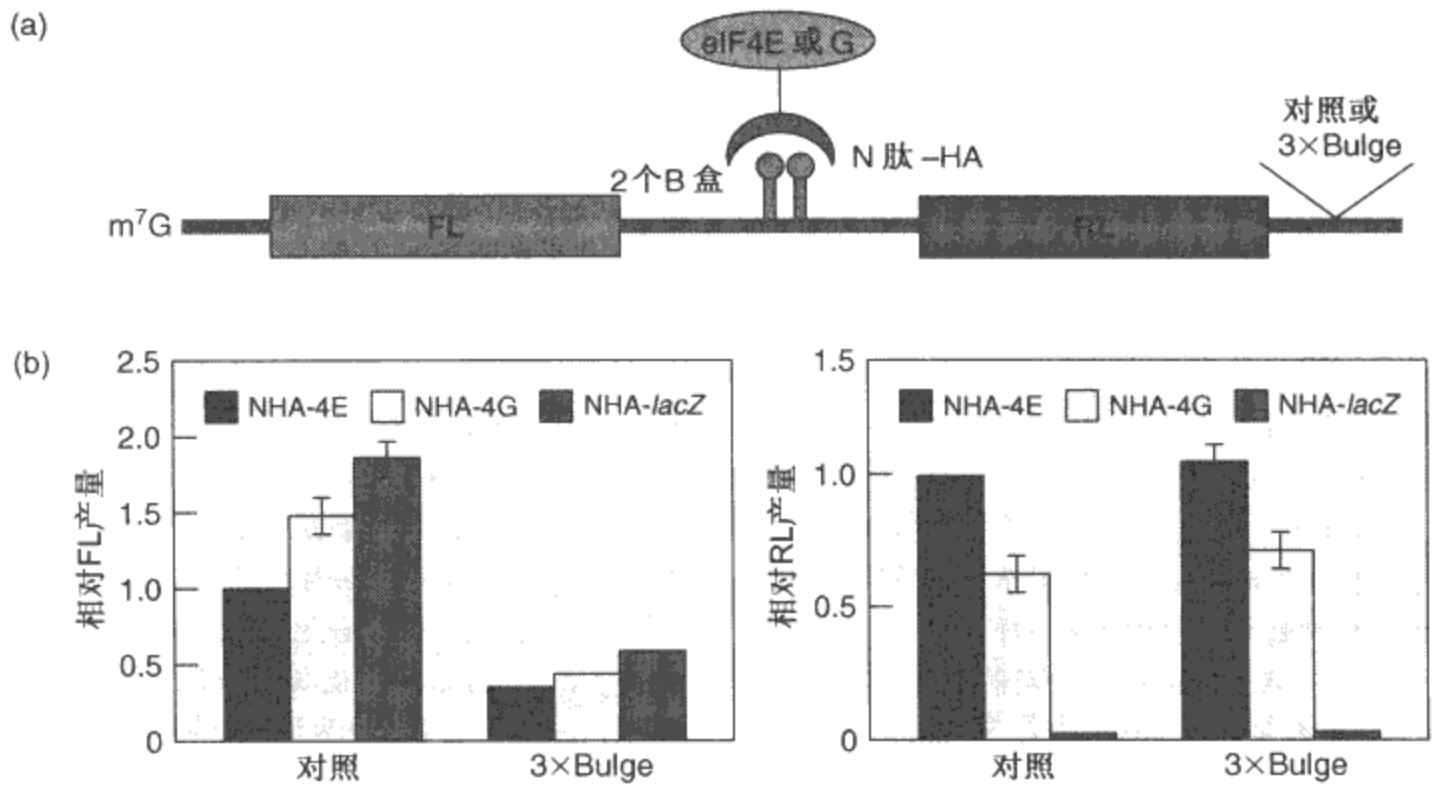


图 17.47　翻译起始因子束缚于双顺反子 mRNA 的顺反子间区的效应。（a）具有两个 B 盒茎环的重组子。两个 B 盒茎环位于两个顺反子之间，并结合到还含有 eIF4E 或 eIF4G 的融合蛋白的 N 肽部分。3′-UTR 含有对照 RL 序列（Con）或 3×Bulge 序列。（b）以不同的蛋白质束缚于顺反子间区域，对照和 3×Bulge mRNA 的 FL（左）和 RL（右）的产物，如图底部所示。束缚于顺反子间区域的 N 肽-红细胞凝集素（NHA)-标签蛋白以颜色标注为 eIF4E 蓝色；eIF4G 黄色；*LacZ* 产物红色。（*Source*：Adapted from Ramesh，S.，et al.，2004 Inhibition of translational initiation by let-7 microRNA in human cells. *Science* 309：1575，fig. 2.）

相比，*let-7* miRNA 抑制 3×Bulge mRNA 的 FL 顺反子的翻译。但是，当 eIF4E 或 eIF4G 被束缚于顺反子间区域时，*let-7* miRNA 不能抑制 3×Bulge mRNA 中 RL 顺反子的翻译（而当束缚于顺反子间区的是 *lacZ* 产物而非起始因子时，几乎没有翻译发生，即使是对照 mRNA）。因此，有了 eIF4E 或 eIF4G（本例通过束缚）就可以避免 *let-7* miRNA 对翻译起始的抑制，由此提示 *let-7* 阻断了 eIF4E 召集 eIF4G 到帽子之前的某个步骤。这个 *let-7* 敏感步骤很明显的候选者是 eIF4E 结合帽子这一步。

在哺乳动物细胞中的这些结果表明，*let-7* miRNA对翻译的干扰不同于第 16 章的线虫中 *lin-4* miRNA 不影响其目标 mRNA 多核糖体合成进而不阻断翻译起始的结果。这种不一致性可以解释为是否不同的 miRNA 具有不同的作用模式，或 miRNA 在不同的生物有不同的作用方式，或两种情况都有。

小结　*let-7* miRNA 将在人细胞中靶 mRNA 的多核糖体转换成较小的多核糖体，表明该 miRNA 阻断人细胞中翻译的起始。由于 IRES 或束缚起始因子产生的帽-非依赖性翻译起始不受 *let-7* miRNA 影响，提示该 miRNA 阻断人类细胞中 eIF4E 对目标 mRNA 帽子的结合。

总结

作为蛋白质合成的前奏，有两个事件必须发生：首先，氨酰-tRNA 合成酶将氨基酸连接到其关连 tRNA 上。这一过程是在一个两步反应中非常专一性完成的，以 ATP 衍生的 AMP 对氨基酸的活化开始。其次，在每一轮翻译的末期，核糖体必须解离为亚基。在细菌里，IF1 活跃地促进这种解离，而 IF3 结合到游离的 30S 亚基上，阻止它与 50S 亚基重新结合形成完整核糖体。

原核生物的起始密码子通常是 AUG，也可以是 GUG，或极少是 UUG。起始氨酰-tRNA 是 *N*-甲酰甲硫氨酸-$tRNA_f^{Met}$。因此，*N*-甲酰甲硫氨酸（fMet）是多肽链的第一个氨基酸，但在蛋白质成熟加工过程中它经常被切除。

30S 起始复合物由游离的 30S 核糖体亚基加 mRNA 和 fMet-$tRNA_f^{Met}$组成。30S 原核核糖体亚基与 mRNA 起始位点的结合依赖于被称作 Shine-Dalgarno 序列的短 RNA 序列与 16S rRNA 3′端互补序列碱基配对。这种结合由 IF3 介导，IF1 和 IF2 因子协助。到此时，所有 3 个起始因子都已经附着在 30S 亚基上了。

IF2 是促进 fMet-$tRNA_f^{Met}$结合到 30S 起始复合

物上的主要因子。另外两个起始因子起重要的支持作用。在 IF2 的生理浓度下，GTP 是 IF2 结合所必需的，但在这个过程中 GTP 并不水解。完整的 30S 起始复合物包括一个 30S 核糖体亚基、各一分子的 mRNA、fMet-tRNA$_i^{Met}$、GTP、IF1、IF2 和 IF3。在 50S 亚基加入 30S 复合物形成 70S 起始复合物之后，GTP 被水解。这一 GTP 水解由 IF2 联合 50S 核糖体亚基执行。水解的目的是从复合物上释放 IF2 和 GTP，以便多肽链延伸能够开始。

真核 40S 核糖体亚基与起始 tRNA（tRNA$_i^{Met}$）一起通过与 mRNA 的 5′帽子结合而找到合适的起始密码子，然后扫描下游结构直到在良好序列环境中发现第一个 AUG。最佳序列环境是：在－3 位是嘌呤，＋4 位是 G。有 5%～10%的情况下，核糖体亚基将越过第一个 AUG，继续搜寻更合适的 AUG。有时核糖体明显起始一个上游的 AUG，翻译一个短的可读框，然后继续扫描，在下游的一个 AUG 再起始。这种机制只在上游 ORF 较短时起作用。靠近 mRNA 5′端的二级结构具有正向或负向影响。紧接 AUG 后的发夹结构可以迫使核糖体亚基暂停于 AUG，因而激发起始。在帽子结构与起始位点之间非常稳定的茎环结构能阻止核糖体亚基扫描，因而阻止起始。

真核生物起始因子有下列一般功能：eIF2 参与 Met-tRNA$_i^{Met}$对核糖体的结合；eIF2B 通过 GTP 取代 GDP 而激活 eIF2；eIF3 结合 40S 核糖体亚基，阻止其与 60S 亚基的重新结合，eIF5 促进 60S 核糖体亚基与 48S 复合物（40S 亚基＋ mRNA＋ Met-tRNA$_i^{Met}$）之间的结合。eIF6 结合到 60S 亚基，阻止 60S 亚基与 40S 亚基的再结合。

eIF4F 是一个帽子结合蛋白，由三部分组成：eIF4E 有实际的帽子结合活性，伴随有另两个亚基 eIF4A 和 eIF4G。eIF4A 有 RNA 解旋酶活性，能够打开在真核生物 mRNA 5′前导序列中发现的发夹结构。eIF4A 完成这一功能由另一个因子 eIF4B 的协助，并需要 ATP 提供能量。eIF4G 是一个接头蛋白，能够结合多种其他蛋白质，包括 eIF4E（帽子结合蛋白）、eIF3（40S 核糖体亚基结合蛋白）和 Pab1p［poly（A）结合蛋白］。通过这些蛋白质的相互作用，eIF4G 能够促进 40S 核糖体亚基结合 mRNA，从而激发翻译起始。

核糖体结合因子 eIF3 是个与 eIF4G 结合的五圆凸蛋白，并通过相同位点结合病毒 IRES 的突出部分。这一发现揭示了 IRES 如何替代 eIF4G 召集 eIF3，进而召集 40S 核糖体亚基到 mRNA 上。eIF3 通过阻断 40S 和 60S 这两个核糖体亚基间的关键接触点而阻止 40S—60S 的过早结合。eIF4G，进而是 eIF4E 靠近结合在 40S 核糖体颗粒上的 mRNA 帽子处，所以 eIF4E 处在执行结合帽子这一任务的位置上。

eIF1 和 eIF1A 协同促使稳定的 48S 复合物的形成，这个复合物包括起始因子、Met-tRNA$_i^{Met}$ 及结合在 mRNA 起始密码子上的 40S 核糖体亚基。eIF1 和 eIF1A 表现为解离 40S 核糖体亚基与 mRNA 之间不合适的复合物，促进扫描起始密码子的 48S 稳定复合物的形成。

eIF5B 与原核生物的 IF2 同源。eIF5B 在结合 GTP 和激发两个核糖体亚基的结合方面与 IF2 相似。在这个反应中，eIF5B 与 eIF5 共同作用。eIF5B 在利用 GTP 水解来促进自身与核糖体分离，使蛋白质合成开始方面与 IF2 相似。与 IF2 不同之处在于它不能促使氨酰-tRNA 结合到小核糖体亚基上。在真核细胞中这一任务由 eIF2 完成。

原核生物的 mRNA 非常短命，所以翻译调控在这类有机体不普遍。然而，一些翻译调控确实发生了。信使 RNA 的二级结构能够调控翻译的起始，如噬菌体 MS2 家族的复制酶基因，或在*E. coli* σ^{32}的 mRNA 中，其翻译起始被二级结构抑制，该二级结构能通过加热而变得松弛。另外，核糖体蛋白能反馈抑制它们自身 mRNA 的翻译。

小 RNA 与蛋白质一致，可以影响 mRNA 的二级结构，控制翻译的起始。核开关是这一控制能够实施的一种方式。*E. coli* 的 *thiM* mRNA 的 5′-UTR 含有一个核开关，包括一个核酸适体。核酸适体可结合硫胺及其代谢物硫胺磷酸，尤其是硫胺素焦磷酸（TPP）。TPP 丰富时，可与核酸适体结合引起 mRNA 构象改变，将 Shine-Dalgarno 序列隐藏在二级结构中。这种转换使 SD 序列不为核糖体所发现，从而抑制 mRNA 的翻译。

真核生物 mRNA 的寿命相对较长，所以对翻译进行调控的机会比原核生物要多。eIF2 的 α 亚基是翻译调控的偏爱靶位。在血红素缺乏的网织红细胞中，HCR 是激活的，因此它能使 eIF2α 磷酸化，抑制翻译起始。在病毒感染的细胞中，另一种激酶 DAI 是活化的，它也使 eIF2α 磷酸化，抑制翻译起始。

胰岛素和若干生长因子可激活由蛋白激酶 mTOR 参与的途径。mTOR 激酶的靶蛋白之一是 PHAS-I（大鼠）或 4E-BP1（人）。通过 mTOR 的磷酸化作用，该蛋白质从 eIF4E 上解离下来，释放 eIF4E 参与更活跃的翻译起始。

在爪蟾卵母细胞中，Maskin 结合 eIF4E 和结合了休眠的细胞周期蛋白 B mRNA 的 CPEB。由于 Maskin 的结合，eIF4E 不能与 eIF4G 结合，所以翻

译受到抑制。在卵母细胞活化时，CPEB 被磷酸化，由此激活多聚腺苷酸化，有可能引起 CPEB 从 Maskin 上解离，Maskin 随后又从 eIF4E 上解离。因为 Maskin 不再黏附，eIF4E 就可以自由地与 eIF4G 结合，翻译能够起始。

铁蛋白 mRNA 的翻译受铁离子诱导。诱导机制大致为：阻遏蛋白（顺乌头酸酶脱辅基蛋白）结合到铁蛋白 mRNA 5′-UTR 区 5′端附近的茎环结构的铁离子应答元件上（IRE）。铁离子解除阻遏并使得 mRNA 翻译得以继续。

let-7 miRNA 将人细胞中靶 mRNA 的多核糖体转换成较小的多核糖体，表明该 miRNA 阻断人细胞中翻译的起始。由于 IRES 或束缚起始因子产生的帽非依赖性翻译起始不受 *let-7* miRNA影响，提示该 miRNA 阻断人类细胞中 eIF4E 对目标mRNA 帽子的结合。

复习题

1. 叙述实验并给出结果，说明核糖体亚基的解离与再结合。

2. IF1 和 IF3 是怎样参与核糖体解离的？绘图并给出与 IF3 相关的证据。

3. 如何知道 IF3 与核糖体 30S 亚基结合，但不能在 50S 亚基加入形成 70S 核糖体后仍保持结合？

4. 两个氨酰-tRNA 叫什么？它们有什么作用？

5. 为什么 MS2 噬菌体复制酶顺反子的翻译依赖于外壳顺反子的翻译？

6. 用结果（无需准确的碱基序列）支持 Shine-Dalgarno 序列和 16S rRNA 碱基配对在翻译起始中的重要作用，选择最令人信服的结果。

7. 给出实验结果，表明 3 个起始因子在 mRNA-核糖体结合中的作用。

8. 叙述实验并给出结果，说明 GTP 水解在 30S 起始复合物形成中的作用（如果有的话）。

9. 叙述实验并给出结果，说明 GTP 水解在 IF2 从核糖体上释放中的作用。

10. 叙述实验并给出结果，说明起始因子对 30S 亚基的协同性结合。

11. 给出实验结果，表明 3 种起始因子在 fMet-$tRNA_i^{Met}$结合核糖体方面的作用。

12. 图示总结*E. coli*. 翻译的起始过程。

13. 解释什么是 Shine-Dalgarno 序列和 Kozak 共同序列，对比它们的作用。

14. 写出理想的真核生物翻译起始位点序列。除 AUG 外，最重要的位点是什么？

15. 图示翻译起始的扫描模型。

16. 举例说明核糖体的扫描能够越过 AUG 并且从下游的 AUG 起始翻译。

17. 在何种情况下，良好序列环境的上游 AUG 对下游 AUG 的起始无阻碍作用，给出证据。

18. 叙述实验并给出结果，说明 mRNA 前导序列的二级结构对扫描的作用。

19. 图示真核生物翻译起始步骤，给出每一级起始因子的作用。

20. 叙述实验并给出结果，鉴定帽子结合蛋白。

21. 叙述实验并给出结果，说明帽子结合蛋白刺激加帽 mRNA，而不是非加帽 mRNA 的翻译。

22. eIF4F 的亚基结构是什么？不要求分子质量。

23. 叙述实验并给出结果，说明 eIF4A 和 eIF4B 在翻译中的作用。

24. 脊髓灰质炎病毒遗传物质是怎样模仿典型的细胞 mRNA 的？有什么不同？病毒是怎样利用这种差异的？与肝炎病毒 C 比较和对比这一行为。

25. 怎样知道 eIF1 和 eIF1A 不能通过激活对同一 mRNA 的扫描，引起复合物 I 转化成复合物 II。

26. 比较起始因子 IF2 和 eIF5B，它们所具有的相同作用是什么？什么功能是 IF2 具有而 eIF5B 没有的？在真核生物中由什么因子执行这一功能？

27. 阐述 *rpoH* mRNA 感知高温并启动自身翻译的机制，这个模型的证据是什么？

28. 阐述*E. coli thiM* 基因中核开关控制翻译的机制。

29. 阐述 eIF2α 磷酸化的翻译抑制模型。

30. 提供一种模型解释 PHAS-I 磷酸化在翻译效率上所起的作用。

31. 叙述实验并给出结果，说明铁蛋白 mRNA 的 IRE 区对铁蛋白合成中铁离子诱导性的重要性。

32. 提出在哺乳动物细胞中铁蛋白合成的铁离子诱导性的假说。确信你的假说能解释为什么当细胞中铁蛋白基因由强启动子驱动时铁蛋白合成不可诱导。

33. 人类 *let-7* miRNA 是怎样控制其目标基因的表达的？总结这一模型的证据。

分析题

1. 叙述一个趾纹分析实验，涉及在无细胞提取物中（包含翻译所需的全部因子）的 *E. coli* 核糖体亚基和假设的 mRNA。只有 30S 亚基或只有 50S 亚基时，预期得到什么结果？两种亚基都存在并有除亮氨酸之外的所有氨基酸时，预期得到什么结果？亮氨酸在该多肽的第 20 位。

2. 预测下列突变对噬菌体 R17 外壳基因和复制酶基因翻译的影响：

a. 外壳基因起始密码子下游 6 个密码子琥珀型突变（提前形成终止密码子）。

b. 外壳基因起始密码子周围茎环的突变减弱了茎环的碱基配对。

c. 复制酶内部基因的突变导致它与外壳基因起始密码子碱基配对。

3. 你正在研究一个真核生物基因，其翻译通常从 mRNA 的第二个 AUG 开始，在两个密码子 AUG 之间的序列是：CGGAUGCACAGGACAUCCUAUGGAGAUGA，两个 AUG 密码子已加下画线，预测下列突变给 mRNA 翻译带来的影响。

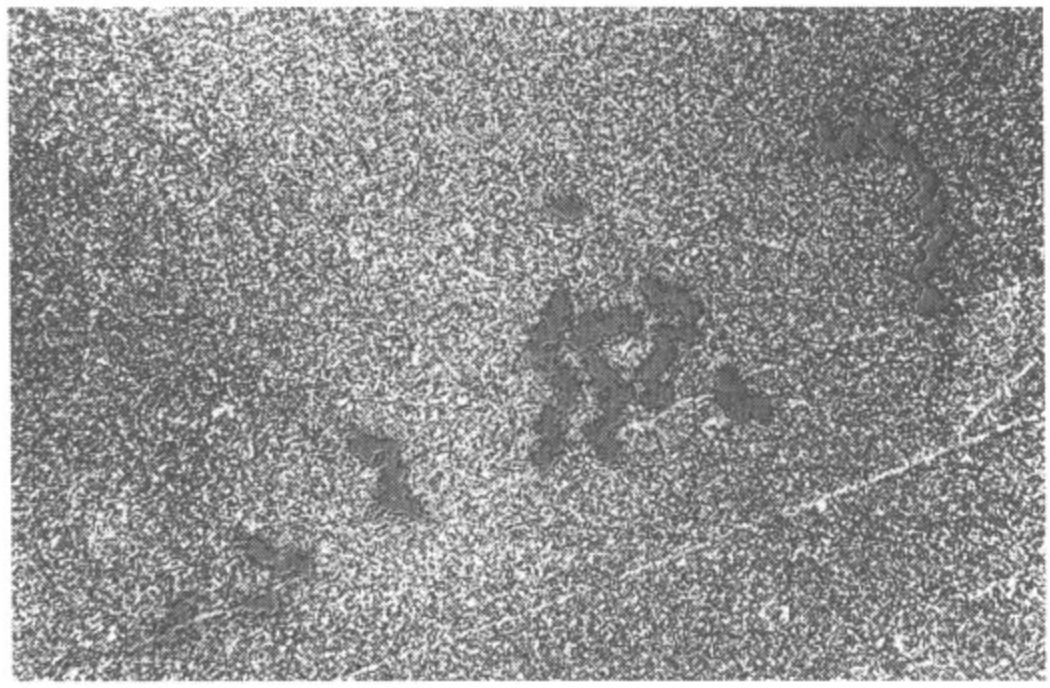
a. 把第一和第二个 C 突变成 G。

b. 把第一和第二个 C 突变成 G，并把第二个 AUG 前的 UAU 密码子突变成 UAG。

c. 把末端的 GAGAUGA 序列突变成 CAGAUGU。

翻译　马　纪　校译　张富春

第 18 章　蛋白质翻译机制 II：延伸与终止



***E. coli* 基因的透射电镜假彩色图像，显示转录和翻译 DNA 纤丝从左下方穿过图像，每个 mRNA 链上黏附有大量核糖体。** *Copyright © Oscar Miller/SPL/Photo Researchers, Inc.*

蛋白质翻译的延伸过程在细菌和真核生物非常类似，所以我们将同时学习这一过程，先讨论细菌系统，然后了解其与真核系统的某些差异。

如第 17 章所学，细菌翻译的起始过程产生了一个由 mRNA 和起始氨基酰-tRNA、fMet-tRNA$_f^{Met}$ 引发的核糖体，多肽链的延伸已经准备好。在我们讨论延伸过程之前，先来考虑一些关于延伸的基本问题：①多肽链是按什么方向合成的？②核糖体按什么方向阅读 mRNA？③指导与 mRNA 对应氨基酸掺入的遗传密码的本质是什么？

18.1　多肽链合成及 mRNA 翻译的方向

蛋白质的合成每次只添加一个氨基酸，那么合成从哪儿开始呢？多肽链的延伸是沿氨基端到羧基端方向，还是沿相反方向？换句话说，哪个氨基酸被第一个添加到延伸的肽链上，是 N 端氨基酸还是 C 端氨基酸？1961 年，Howard Dintzis 提供了多肽链合成沿 N 端→C 端的确凿证据，他在研究分离的兔网织红细胞（未成熟血细胞）中 α-和 β-球蛋白的合成时获得了这一证据。他用 [^{3}H] 亮氨酸间隔不同时间段标记合成中的球蛋白，以 [^{14}C] 亮氨酸进行长时间标记。然后，分离 α-和 β-球蛋白，用胰蛋白酶切成多肽，分离肽段。以肽段在蛋白质中从 N 端到 C 端的位置对 [^{3}H] 亮氨酸掺入多肽中的相对量作图。所用 [^{14}C] 亮氨酸的长时间标记应该对所有多肽是同等的，所以可作为内标对照计算在纯化过程中某些肽的损失，以及不同肽段间亮氨酸含量的差异。

图 18.1 显示了这一过程是如何判定蛋白质翻译方向的。需要注意的是，当添加标记物时，蛋白质链处于合成的各不同阶段。因此，有些刚开始，有些已部分完成，而有些则基本完成了。这意味着只有当标记物加入刚开始合成的蛋白质时，标记物才会掺入第一个氨基酸中，其他不同合成程度的蛋白质将在下游肽段中出现标记物，而不是在第一个氨基酸中。对比之下，短时间标记后蛋白质合成末端处的标记物会相对丰富，中间肽段为中等水平的标记。因此，如果翻译从氨基端起始，标记将在羧基末端最强。图 18.2 显示了这一结果。α-和 β-球蛋白中标记的多肽都自氨基端向羧基端升高，并且这种差异在短时间内特别明显。因此，蛋白质合成起始于氨基端。

mRNA 是按 5′→3′方向阅读，还是相反方向呢？已知蛋白质的合成方向是从氨基端至羧基端的，那么很容易证明 mRNA 是按 5′→3′方向被阅读的。20 世纪 60 年代，当分子生物学家第一次用人工合成的 mRNA 做模板合成蛋白质时，其中一些信使 RNA 就包含了对该问题的答案。例如，当 Ochoa 及同事翻译 mRNA 5′-AUGUUUn-3′时，得到了 fMet-Phe$_n$，其

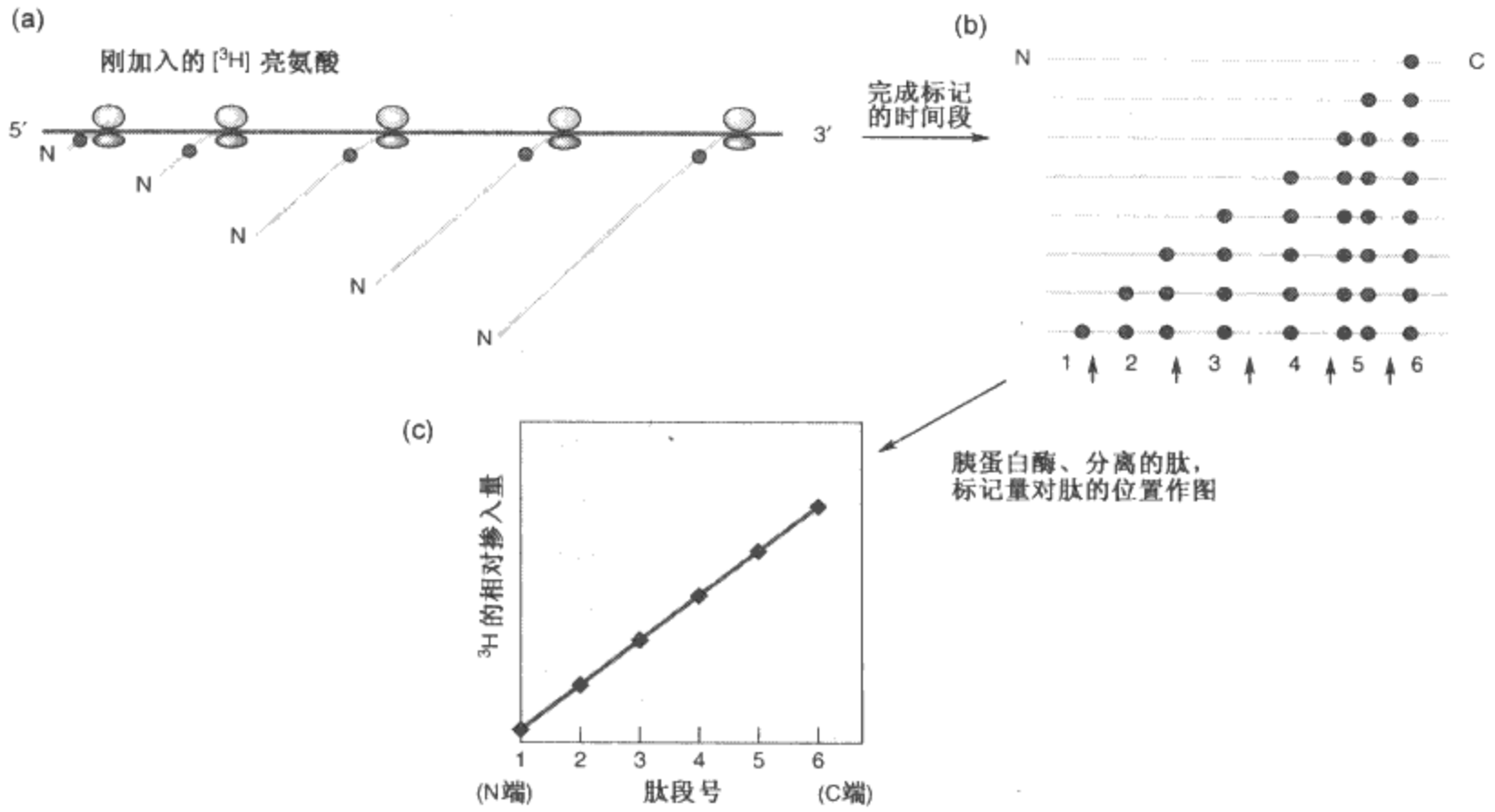


图 18.1　确定蛋白质翻译方向的实验策略。（a）标记蛋白质。多个核糖体在一条 mRNA 链上进行翻译，假设 mRNA 的翻译方向是 5′→3′，蛋白质的合成方向是从氨基端（N）到羧基端（C）。标记的氨基酸（^{3}H-Leu）刚刚被加入系统，所以它已经开始掺入延长中的蛋白质链，如圆点所示。在刚开始合成的蛋白质中，标记靠近左边多肽的 N 端，而在右边几乎完成的蛋白质中，标记靠近多肽的 C 端。（b）一段中等标记时间后，在完整蛋白质中标记的分布。靠近顶部的蛋白质，标记只在 C 端，与（a）图靠近右侧基本上完成的蛋白质相对应；底部的蛋白质，标记向 N 端分布，对应于（a）图左侧正在增长的蛋白质。这些蛋白质有时间将标记掺入不同长度的蛋白质中。胰蛋白酶的切割位点用下方的箭头表示，根据在蛋白质中的位置，产生的肽段标号为 1～6。（c）模式实验结果。将 1～6 号肽段与其^{3}H 标记相对量作图，发现 C 端的多肽标记最强。这正是我们所预期的翻译从 N 端起始的结果。假如从 C 端起始（与 a 图中所画的相反），那么 N 端肽段的标记应该是最强的。

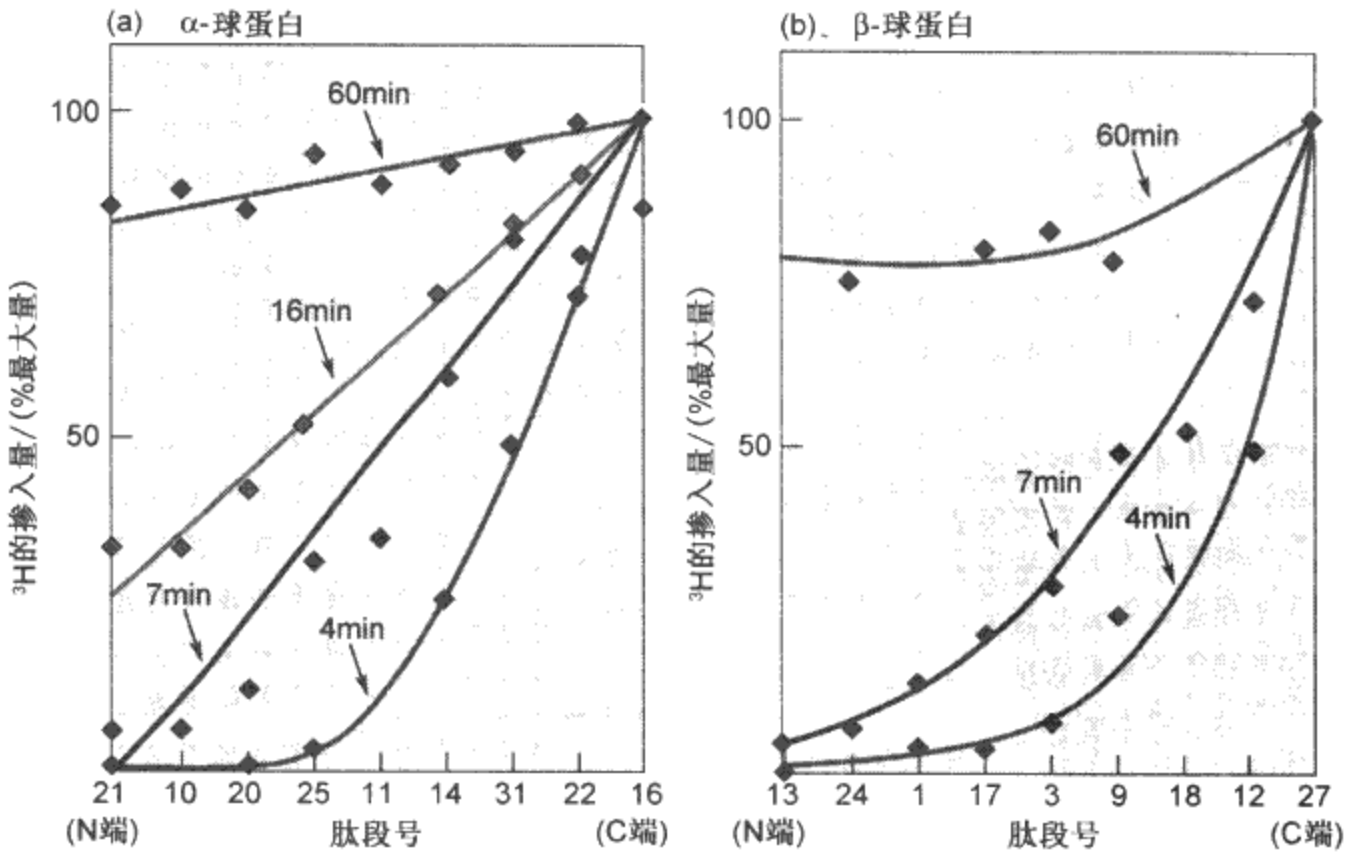


图 18.2　确定翻译方向。Dintzis 在兔网织红细胞中实施了图 18.1 所示的实验计划，该系统只合成 α-，β-球蛋白。Dintzis 用［^{3}H］-Leu 在不同时间段标记网织红细胞，然后分离 α-、β-球蛋白，用胰蛋白酶分别水解，再确定每条肽段的标记量。用^{3}H-Leu 的相对量对肽段数目作图，N 端肽在左侧，C 端肽在右侧。曲线（a）α-球蛋白、（b）β-球蛋白都表明 C 端多肽有最大标记，尤其是短时间标记后。如果翻译从蛋白质的 N 端开始，这正是我们所期望的。注意，肽段的号码并不像图 18.1 中那样与它们在蛋白质中的位置对应。（*Source*：Adapted from Dintzis，H. M.，Assembly of the peptide chains of hemoglobin. *Proceedings of the National Academy of Sciences USA* 47：255，1961.）

中 fMet 是在氨基端的。已知 AUG 编码 fMet，UUU 编码丙氨酸（Phe），我们又知道 fMet 被掺入蛋白质的氨基端，这就意味着它是在任一 Phe 之前先加入的。因此，mRNA 一定是从 5′端阅读的，因为那儿是 fMet 密码子的位置。

小结 信使 RNA 按 5′→3′方向阅读，与它们被合成的方向一致。蛋白质合成以氨基端→羧基端方向进行，这意味着氨基端氨基酸先加入。

18.2 遗传密码子

遗传密码（genetic code）一词是指 mRNA 上编码 20 种氨基酸的 3-碱基密码（**密码子**，codon）。像其他密码一样，在我们了解密码子所代表的意思之前必须破译它。然而在 1960 年之前，有关遗传密码的一些基础问题尚不清楚，如密码子重叠吗？密码中有间隔或间断吗？几个碱基组成一个密码子？在 20 世纪 60 年代，这些问题通过一系列富有想象力的实验都得到了回答，这里我们来了解一下这些实验。

非重叠性密码子

在非重叠密码中，大多数情况下每个碱基都是密码子的一部分。而在重叠密码中，一个碱基可能是两个，甚至 3 个密码子的成分。请看下面的一小段信使 RNA：

AUGUUC

假设密码是三联体（每个密码子有 3 个碱基）并且从起始端读起，那么如果密码是非重叠的，则密码子应该是 AUG 和 UUC。而重叠密码可能会产生 4 个密码子：AUG、UGU、GUU、UUC。早在 1957 年，Sydney Brenner 从理论上推测这种完全重叠的三联密码是不可能的。

然而截至 1957 年，密码子部分重叠的可能性仍然不能排除。但 A. Tsugita 等通过以下分析否定了这一可能：如果密码子是非重叠性的，那么 mRNA 的一个碱基改变（错义突变）仅会引起所得蛋白质中一个氨基酸的改变。如下面的序列：

AUGCUA

假设密码为三联体（每个密码子 3 个碱基）并且从起始端读起，那么如果是非重叠的，则密码应该是 AUG 和 CUA。第三个碱基（G）的改变只会改变一个密码子（AUG），因而最多影响一个氨基酸。相反，如果密码是重叠的，碱基 G 是 3 个相邻密码子（AUG、UGC 和 GCU）的一部分，那么改变 G 就会产生 3 个氨基酸的改变。当研究者将单碱基改变引入烟草花叶病毒（TMV）mRNA 时，发现并未引起一个以上氨基酸的改变。因此，密码子一定是非重叠性的。

密码中无间隔

如果密码中包含非翻译间隔，或“间隔符”，那么单碱基缺失或增加就会改变多个密码子，但预期在下一个间隔符之后，核糖体会回到正常轨道。换句话说，这些突变可能经常是致死性的，但在多数情况下，只在信使 RNA 的一个间隔符前发生，因此，如果有影响的话也会很小。但是，如果没有间隔符使核糖体回到正常轨道，那么突变将是致死性的，除非突变刚好发生在信使 RNA 的末端。

这类突变确实可以发生，被称为**移码突变**（frameshift mutation），其机制如下。再看以下这一小段信使 RNA：

AUGCAGCCAACG

如果翻译从头开始，密码子应该是 AUG、CAG、CCA、ACG。如果紧接在 U 的后面插入一个碱基（X），得到：

AUXGCAGCCAACG

现在，这个序列翻译为 AUX、GCA、GCC、AAC。注意，插入的额外碱基不但改变了它所插入的密码子（AUX），而且影响了其后的所有密码子。**可读框**（reading frame）向左偏移了一个碱基，因此在第二个密码子中原来 C 是第一个碱基，现在是 G 在那个位置。

另一方面，带有间隔符的密码的两侧会有一个或多个非翻译碱基，在下列信息中用 Z 表示。间隔符起到分隔密码子的作用，以便核糖体可以识别它：

AUGZCAGZCCAZACGZ

在这个序列的任意位置切除或插入一个碱基只会改变一个密码子，受损密码子末端的间隔符（Z）会使核糖体返回到正常翻译轨道。因此，对第一个密码子加上一个额外碱基（X）产生下列信息：

AUXGZCAGZCCAZACGZ

现在，第一个密码子（AUXG）是错的，但其他密码子仍由间隔符 Z 规则地隔开并正常翻译。

但是，当 Francis Crick 及同事用吖啶染料处理细菌（该染料通常引起单碱基插入或缺失）时，他们发现这些突变的后果非常严重，突变基因都产生没有功能的产物。这正是我们所预期的“无间隔符”密码没有间隔的结果：碱基的插入或缺失造成信使 RNA 的阅读框变化，这种改变一直持续到信息的末端。

此外，Crick 发现插入一个碱基可以抵消缺失碱基所造成的影响，反之亦然。这一现象如图 18.3 所示，先人工合成一段由密码子 CAT 重复组成的基

因，当在第三位插入碱基 G 后，可读框发生改变，其后所有密码子都读成 TCA 了。当以野生型基因开始，删除第 5 位碱基 A 时，可读框从另一个方向发生改变，其后的所有密码子都读成 ATC 了。将这两个突变基因杂交可产生一个重组的“假野生型”基因（图 18.3 中第 4 行），其中前两个密码子 CAG 和 TCT 是错的，但其后的插入和切除效应抵消，原初的可读框恢复，此后的所有密码子读为 CAT。

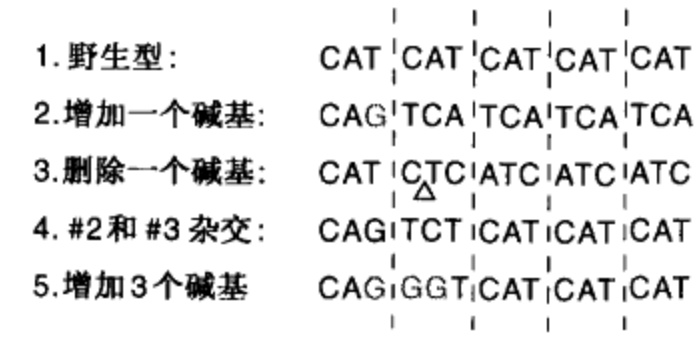


图 18.3　移码突变。第 1 行：由 CAT 重复序列组成的人造基因。垂直虚线表示从起始密码子开始的可读框。第 2 行：在第 3 位加入碱基 G，使第一个密码子变为 CAG，可读框依次向左移动一位导致其后的密码子全部变为 TCA。第 3 行：删除野生型基因的第 5 位碱基 A（三角形标记），第 2 个密码子变为 CTC，可读框依次向右移动一位，其后的所有密码子变为 ATC。第 4 行：将第 2 行和第 3 行的突变基因杂交可获得“假野生型”回复突变体，其密码子顺序因添加和删除效应回复，最后的结果是只有开始的两个密码子改变，而其他的又回到正确的可读框。第 5 行：在前两个碱基后加入 3 个碱基 GGG（红色），打乱了前两个密码子，但其后阅读框的顺序未发生改变，缺失 3 个碱基也得到同样结果。

三联密码

Francis Crick 和 Leslie Barnett 发现，一组假定的 3 个碱基插入或缺失会产生假野生型基因（图 18.3，第 5 行）。这必然要求密码子由 3 个碱基组成。正如 Crick 在看到实验结果后对 Barnett 说的“我们是唯一两个知道密码是三联体的人!”实际上，Crick 和 Barnett 只是推断出他们的假野生型基因含有 3 个碱基插入或缺失，他们尚无法测序基因加以证实，因此还需要更多的实验来证实。

1961 年，Marshall Nirenberg 和 Johann Heinrich Matthaei 做了一个突破性的实验，为证明密码子的三联体特性以及为破译遗传密码奠定了基础。这个实验难以置信的简单，它表明人工合成的 RNA 可以进行体外翻译。具体讲，当 Nirenberg 和 Matthaei 翻译 poly（U）（只有 U 组成的人工合成 RNA）时，得到了多聚苯丙氨酸。这一结果自然告诉他们编码苯丙氨酸的密码子只含 U，这个发现本身就很重要，但其更大的提示是可以设计确定序列的人工合成 mRNA，分析其蛋白质产物从而使密码的性质清楚地显示出来。Gobind Khorana 及同事是这个策略的主要实践者。

以下是 Khorana 如何合成信使 RNA 证实密码子含有 3 个碱基的。首先，如果密码子含有奇数个碱基，那么重复的二核苷酸 poly（UC）或 UCUCUCUC…，不管翻译起始于哪个碱基，都应当包含两个交替的密码子（UCU 和 CUC）。其产生的蛋白质应该是重复的二肽（dipeptide），即两个氨基酸相互交替。如果密码子有偶数个碱基，则只有一个密码子（如 UCUC）重复下去。当然，如果翻译起始于第二个碱基，单个重复密码子会不同（CUCU）。在这两种情况下，所产生的蛋白质应是同聚多肽（homopolypeptide），只含有一种氨基酸。Khorana 发现 poly（UC）翻译成一条重复的二肽 poly（Ser-Leu）（图 18.4a），证明密码子含有奇数个碱基。

正如所预期的，如果密码子中碱基数是 3 或 3 的倍数，重复三联码就被翻译成同聚多肽。例如，poly（UUC）翻译成多聚苯丙氨酸加多聚丝氨酸加多聚亮氨酸（图 18.4b）。产生 3 种不同产物的原因是翻译可从人工合成信息的任一点开始。因此，poly（UUC）可读作 UUC、UUC…，UCU、UCU…，或 CUU、CUU…，这取决于翻译从哪里开始。在所有情况下，一旦翻译开始，只要密码子中碱基数能被 3 整除，就只有一种密码子。

重复的 4 核苷酸被翻译成重复的四肽。例如，poly（UAUC）产生多聚（酪氨酸-亮氨酸-丝氨酸-异亮氨酸）（图 18.4c）。作为练习，你可以依据 3 碱基、9 碱基或更多碱基密码自己写出这类信使 RNA 的氨基酸序列，但不必要做 6 碱基密码子（我们已经知道密码子不可能是偶数）。由于密码子不太可能有 9 个碱基之多，所以 3 是最佳选择。以另一种方式来看这个问题：3 是给出足够的不同密码子对 20 种氨基酸进行专一性编码的最小数字（4 个碱基取 3 的排列数为 4^3，即 64）。2 碱基密码子只有 16 个（$4^2=16$），不够用，而 9 碱基密码子超过了 200 000 个（$4^9=262\,144$），大自然通常比这要节约得多。

小结　遗传密码是 mRNA 上的一套 3 碱基密码，指导核糖体将特定的氨基酸加入多肽。这种密码是非重叠性的，即每个碱基只是一个密码子的一部分，密码间也缺乏间隔或间隔符，即在 mRNA 编码区的每个碱基是一个密码子的一部分。

(a) UCUCUCUCUCUC
Ser Leu Ser Leu

(b) UUCUUCUUCUUC 或 UUCUUCUUCUUC 或 UUCUUCUUCUUC
Phe Phe Phe Phe　　Ser Ser Ser　　Leu Leu Leu

(c) UAUCUAUCUAUC
Tyr Leu Ser Ile

图 18.4　几种合成 mRNA 的编码性质。(a) 由 UC 二核苷酸重复片段构成的 RNA 链包含两个交替的密码子：UCU 和 CUC，分别编码 Ser 和 Leu。翻译产物是 Ser-Leu 二肽重复序列组成的多聚体。(b) 由 UUC 三联体重复序列构成的 RNA 链包含 3 个密码子：UUC、UCU 和 CUU，分别编码 Phe、Ser 和 Leu。根据翻译的起点不同，即核糖体上的可读框不同分别得到 3 种产物：多聚苯丙氨酸、多聚丝氨酸和多聚亮氨酸。(c) 由 UAUC 重复片段构成的 RNA 链包含 4 个密码子：UAU、CUA、UCU 和 AUC，分别编码 Tyr、Leu、Ser 和 Ile。翻译产物是由 Tyr-Leu-Ser-Ile 四肽重复序列组成的多聚体。

破译密码

显然，Khorana 的人工合成 mRNA 提供了某些密码子的重要线索。例如，poly（UC）产生多聚丝氨酰-亮氨酸，据此可知两个密码子（UCU 和 CUC）中的一个编码 Ser，另一个编码 Leu。但问题依然存在：哪个密码子编码 Ser，哪个编码 Leu? Nirenberg 设计了一套非常有效的方法来回答这个问题。他发现三核苷酸通常就足以像 mRNA 了，可引起专一的氨基酰-tRNA 与核糖体结合。例如，三核苷酸 UUU 可引起苯丙酰-tRNA，而不是赖氨酰-tRNA或其他任何氨酰-tRNA 的结合，因此可认为 UUU 是苯丙氨酸的密码子。但这个方法还不完美，因为有些密码子不能引起任何氨酰-tRNA 与核糖体结合，虽然它们是某些氨基酸真正的密码子。但该方法对 Khorana 的方法提供了很好的补充，后者自身也不能至少是不容易得出全部的答案。

这是两种方法如何结合使用的例子：多聚核苷酸 poly（AAG）产生多聚赖氨酸加多聚谷氨酸加多聚精氨酸，这段合成的 RNA 链有 3 个不同的密码子：AAG、AGA 和 GAA。哪一个编码赖氨酸呢? 用 Nirenberg 的分析法对所有 3 个都进行了测试，得到的结果见图 18.5。很明显，AGA 和 GAA 不能引起 [^{14}C] 赖氨酰-tRNA 与核糖体结合，而 AAG 可以。因此，在 poly（AAG）中，AAG 是 Lys 的密码子。对这个实验还值得注意的是 AAA 也引起赖氨酰-tRNA 的结合，因此 AAA 是赖氨酸的另一个密码子，由此表明该密码系统的一般特性：多数情况下，有一个以上的三联体编码一个氨基酸。换句话说，密码是**简并的**（degenerate）。

图 18.6 给出了全部遗传密码。正如预计的，有

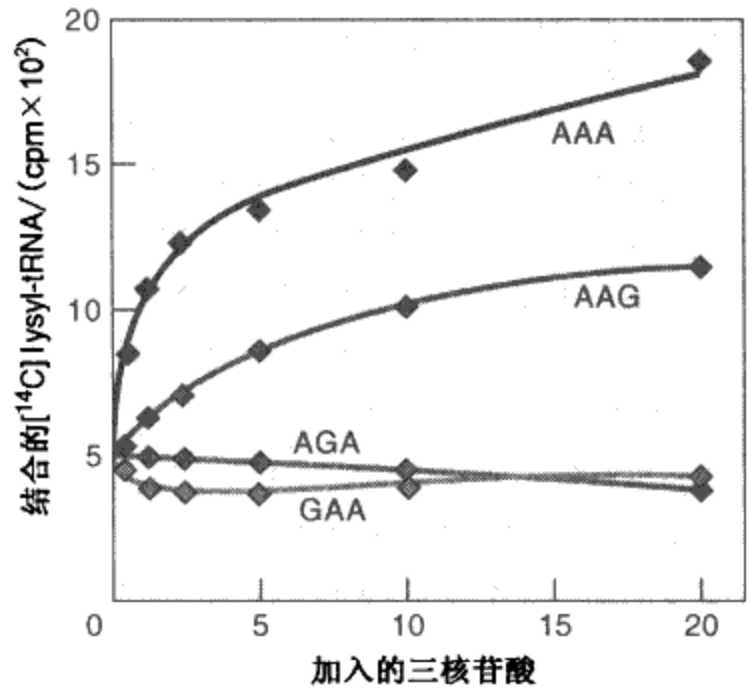


图 18.5　不同密码子所引发的赖氨酰-tRNA 与核糖体的结合。将同位素碳标记（^{14}C）的赖氨酰-tRNA，在下列三核苷酸 AAA、AAG、AGA、GAA 存在情况下与*E. coli*的核糖体混合。用硝酸纤维素薄膜过滤法检测赖氨酰-tRNA-核糖体复合物的形成（复合物会黏附在薄膜上，而没有形成复合物的赖氨酰-tRNA 不发生黏附）。AAA 是已知的赖氨酸密码子，因此，也同样可以观察到由它引发的赖氨酰-tRNA 与核糖体结合的现象。（*Source*：Adapted from Khorana, H. G., Synthesis in the study of nucleic acids, *Biochemical Journal* 109：715，1968.）

64 个不同的密码子但只有 20 种氨基酸，并且所有的密码子都被用上了。有 3 个“终止”密码子出现在信使 RNA 的末端，其他所有的都编码特定的氨基酸，这意味着该套密码是高度简并的。亮氨酸、丝氨酸、精氨酸各有 6 个不同的密码子，脯氨酸、

苏氨酸、丙氨酸各有 4 个，异亮氨酸有 3 个，其余的大多数氨基酸有 2 个。只有甲硫氨酸和色氨酸各只有一个密码子。

小结 遗传密码的破译是通过人工合成的信使 RNA 来观察合成的多肽链，或者人工合成三核苷酸观察氨酰-tRNA 与核糖体的结合进行的。共有 64 个密码子，3 个为终止信号，其他密码子都编码氨基酸，这意味着密码子是高度简并的。

密码子与反密码子间的异常碱基配对

生物体如何解决多个密码子编码同一氨基酸这个问题呢？一种方式可能是对同一个氨基酸具有多个 tRNA（**同工 tRNA**，isoaccepting species），每个都专一性地对应不同的密码子。这是答案的一部分，事实上，特定生物体中的确含有约 60 种不同的 tRNA。但原则上讲，比这个简单的假设所预计的要少得多的 tRNA 就可以解决问题了。Francis Crick 又以其卓有远见的理论预见了实验结果。他假设密码子的前两位碱基必须按照 Watson-Crick 碱基配对原则与反密码子严格配对（图 18.7a），但是最后一位碱基则会从其正常位置发生“摇摆”与反密码子形成非正常配对，这个假说被称为**摇摆假说**（wobble hypothesis）。具体讲，Crick 认为反密码子中的一个 G 不仅能与密码子最后一位（摇摆位）的 C 配对，而且也能与 U 配对，由此产生**摇摆碱基配对**（wobble base pair），如图 18.7b 所示。注意 U 是如何从其正常位置摇摆而形成这个碱基对的。

进一步，Crick 注意到 tRNA 上有一个特殊的**次黄嘌呤核苷**（inosine，I），其结构类似于 G，可正常地像 G 一样配对，预计能与 C 配对（Watson-Crick 配对）或在第三位（摇摆位）与 U 配对（摇摆配对）。但 Crick 认为次黄嘌呤核苷还可以形成另一种摇摆配对，即与密码子第三位的 A（图 18.7c）配对。这就意味着以 I 在第一位的反密码子可以和第三位是 C、U 或 A 的 3 个不同密码子配对。

摇摆现象降低了翻译遗传密码所需的 tRNA 数目。例如，根据摇摆假说，编码 Phe 的 2 个密码子 UUU 和 UUC（图 18.6 左上方），都可以被反密码子 3′-AAG-5′（图 18.8a）识别。反密码子 5′端的 G 可以与 UUC 的 C 形成 Watson-Crick 配对，或与 UUU 的 U 发生 G-U 摇摆配对。同样，亮氨酸密码子 UUA 和 UUG 都可以被反密码子 3′-AAU-5′（图 18.8b）识别，反密码子 5′端的 U 可以与 UUA 的 A 进行 Watson-Crick 配对，或与 UUG 的 G 摇摆配对。

第二个位置

第一个位置(5′端)	U	C	A	G	第一个位置(3′端)
U	UUU Phe UUC Phe	UCU Ser UCC Ser UCA Ser UCG Ser	UAU Tyr UAC Tyr	UGU Cys UGC Cys	U C
	UUA Leu UUG Leu		UAA STOP UAG STOP	UGA STOP UGG Trp	A G
C	CUU Leu CUC Leu CUA Leu CUG Leu	CCU Pro CCC Pro CCA Pro CCG Pro	CAU His CAC His	CGU Arg CGC Arg CGA Arg CGG Arg	U C A G
			CAA Gln CAG Gln		
A	AUU Ile AUC Ile AUA Ile	ACU Thr ACC Thr ACA Thr ACG Thr	AAU Asn AAC Asn	AGU Ser AGC Ser	U C
	AUG Met		AAA Lys AAG Lys	AGA Arg AGG Arg	A G
G	GUU Val GUC Val GUA Val GUG Val	GCU Ala GCC Ala GCA Ala GCG Ala	GAU Asp GAC Asp	GGU Gly GGC Gly GGA Gly GGG Gly	U C A G
			GAA Glu GAG Glu		

图 18.6 遗传密码。列出了全部 64 个密码子以及所编码的氨基酸。注意，3 个红色的是终止信号。

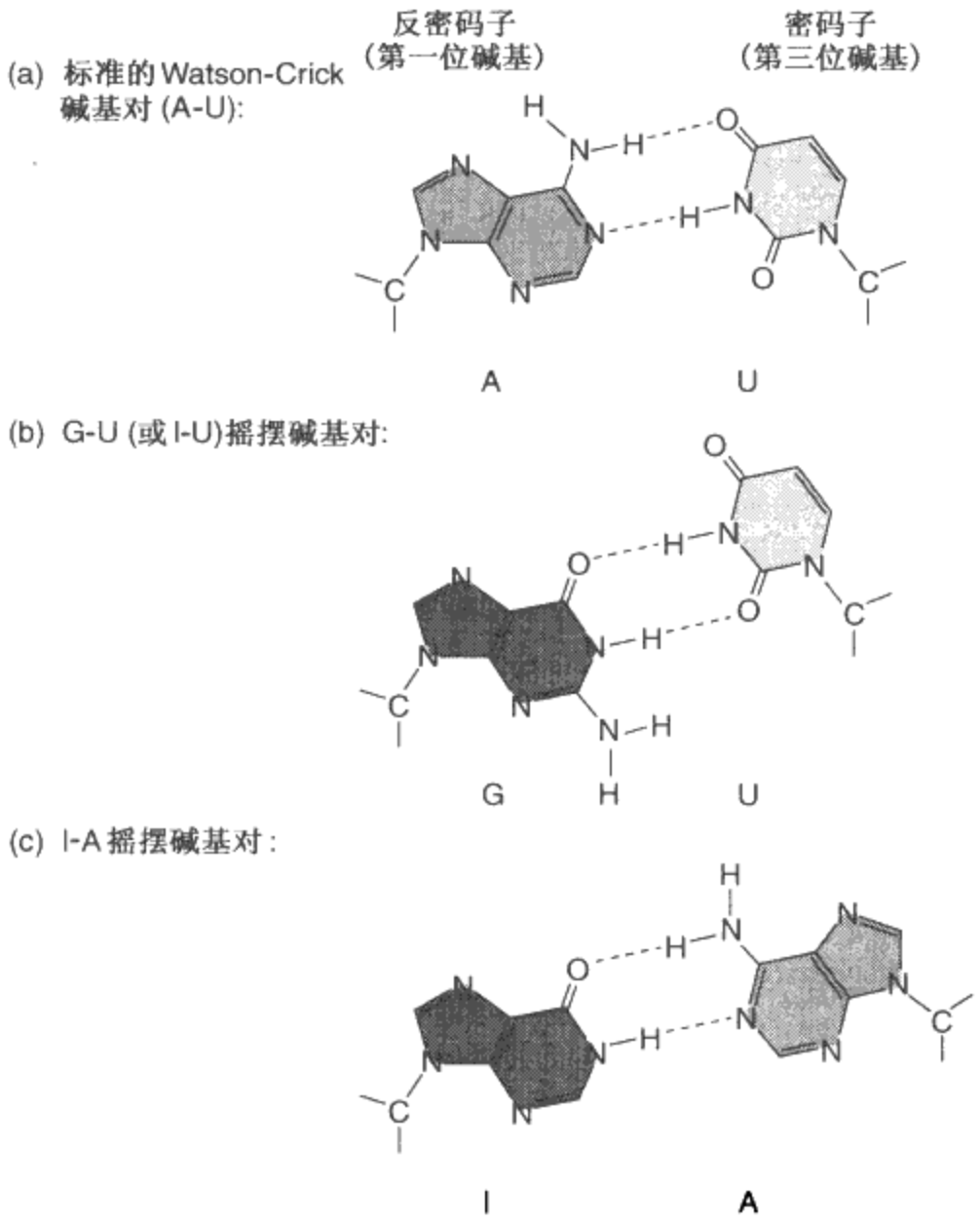


图 18.7　摇摆碱基配对。（a）在标准（A-U）碱基对中碱基的相对位置。这里左边的以及在摇摆碱基对（b）和（c）中的碱基是反密码子的第一位碱基。右侧的碱基是密码子的第三位碱基。（b）在G-U（或 I-U）摇摆碱基对中碱基的相对位置。注意，U 必须向上“摇摆”才能与 G（或 I）配对。（c）在 I-A 摇摆碱基对中碱基的相对位置。A 必须逆时针摆动才能与 I 形成配对。

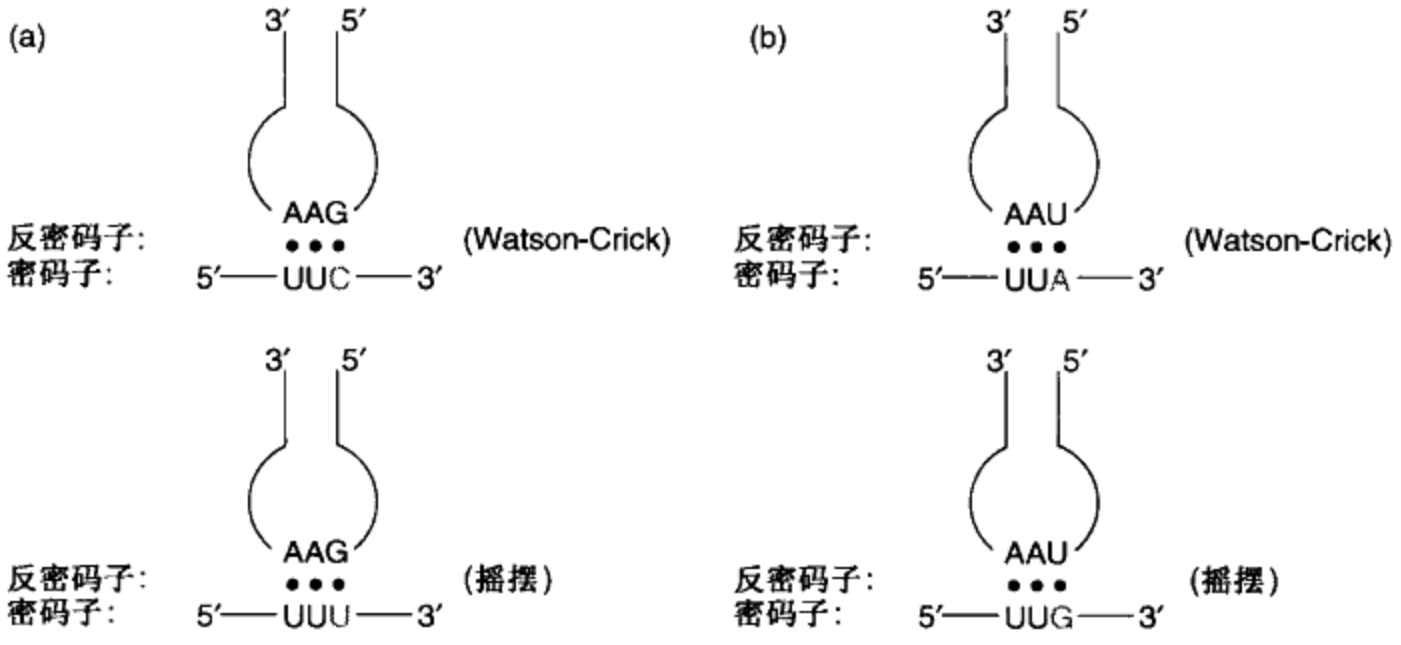


图 18.8　摇摆的位置。（a）缩写的 tRNA 带有反密码子 3′-AAG-5′与编码 Phe 的两个密码子 UUC 和 UUU 碱基配对。摇摆部位（密码子的第三位碱基）用红色标记。与 UUC 的碱基配对（上方）只采用 Watson-Crick 配对；与 UUU 的碱基配对（下方），在密码子的前两位采用两个 Watson-Crick 配对，但在摆动位要求摇摆配对（G-U）。（b）同样，含有反密码子 3′-AAU-5′的 tRNA 与编码 Leu 的两个密码子 UUA 和 UUG 配对。与 UUG 密码子配对时，要求摇摆部位发生 G-U 摆动配对。

小结　遗传密码简并性的一部分是由同工 tRNA提供的，同工 tRNA 结合同一个氨基酸但识别不同的密码子，其余部分则由摇摆来决定，其中密码子第三位碱基被允许稍微偏离其正常位置，

与反密码子形成非 Watson-Crick 配对，由此允许同一个氨酰-tRNA 与一个以上密码子配对。摇摆碱基对包括 G-U（或 I-U）和 I-A。

（几乎）通用的密码

在遗传密码破译后的几年中，从细菌到人类，验证过的所有生物都共用这套相同密码。因此，普遍认为遗传密码是通用的，没有任何变化（随后我们将看到这一观点是不正确的）。这种显而易见的普遍性引出了地球生命的单一起源论。

产生这一观点的原因是：我们看见的每个特定密码子的分配没有什么遗传上的优势。例如，没有明显的理由解释为什么 UUC 对苯丙氨酸是个好密码子，而 AAG 对赖氨酸是好密码子。相反，遗传密码也许是个“偶然事件”，它只是碰巧这样进化的。然而，一旦这些密码子形成，就有很好的理由解释为什么不再发生变化了，因为发生根本性改变几乎一定是致命的。

例如，对于半胱氨酸的 tRNA 及其识别的密码子 UGU，要使二者关系改变，那么半胱氨酰-tRNA 的反密码子必须改变以便识别不同的密码子，如丝氨酸密码子 UCU。同时，在该生物基因组中编码重要丝氨酸的所有 UCU 密码子都必须变成另一个丝氨酸密码子，以便它们不被认做半胱氨酸密码子。即使在漫长的进化过程中，所有这些事情同时发生的机会也是可以忽略不计的。这就是为什么遗传密码有时又被称为是“冻结的偶然事件”：密码子一旦形成，不管什么理由，都得保持不变。因此，通用密码是生命同一起源的强有力证据。毕竟，如果生命独立起源于两个地方，很难期望两个物种碰巧进化出相同的遗传密码！

对于以上讨论值得注意的是，遗传密码并不是绝对通用的，也存在一些例外。这些例外首先是在线粒体基因组中发现的，在果蝇线粒体中，UGA 是色氨酸密码子，而不是“终止”信号。更明显的是，AGA 在线粒体编码丝氨酸，而在标准编码中是精氨酸密码子。哺乳动物线粒体也表现出一些差异，AGA 和 AGG 虽然在标准编码中都是精氨酸密码子，但在人和牛的线粒体基因组中意义却不同，它们编码“终止子”。此外，AUA 通常是异亮氨酸密码子，而在人和牛线粒体中编码甲硫氨酸。

由于这些异常发生在线粒体中，显得相对不重要而被忽略。线粒体基因组非常小，只编码几种蛋白质，因此比核基因组有更多的自由发生变化。但是，在核基因组和原核生物基因组也发现了例外的密码子。在至少 3 种纤毛原生动物中，包括草履虫（*Paramecium*），通常是终止密码子的 UAA 和 UAG 编码谷氨酰胺。在原核生物支原体（*Mycoplasma capricolum*）中，正常是终止密码子的 UGA 编码色氨酸。在病原性酵母（*Candida albicans*）中，通常的亮氨酸密码子 CTG 编码丝氨酸。偏离标准遗传密码的情况见表 18.1。

显然，所谓的通用密码并不是真正通用的，这是否意味着地球生命是多起源的呢？如果异常密码与标准密码根本上不同，这种可能性也许很大，但它们不是。在多数例子里，异常密码子是终止密码子被用于编码谷氨酰胺或色氨酸。关于这类现象有一

表 18.1　对“通用”遗传密码的变异

来源	密码子	通常的含义	新的含义
果蝇线粒体	UGA	终止	色氨酸
	AGA 和 AGG	精氨酸	丝氨酸
	AUA	异亮氨酸	甲硫氨酸
哺乳动物线粒体	AGA 和 AGG	精氨酸	终止
	AUA	异亮氨酸	甲硫氨酸
	UGA	终止	色氨酸
酵母线粒体	CUN	亮氨酸	苏氨酸
	AUA	异亮氨酸	甲硫氨酸
	UGA	终止	色氨酸
高等植物线粒体	UGA	终止	色氨酸
	CGG	精氨酸	色氨酸
Candida albicans 细胞核	CTG	亮氨酸	丝氨酸
原生动物细胞核	UAA 和 UGA	终止	谷氨酰胺
Mycoplasma	UGA	终止	色氨酸

个广为接受的机制，我们将在本章后面谈到。绝大多数已知密码子改变编码氨基酸的例子都发生在线粒体中，而线粒体基因组由于其编码的蛋白质数量远比核基因组甚至原核生物基因组少，因而安全地、时不时地改变一个密码子也是可以预见的。总之，即使密码不完全通用，但标准密码确实存在，并且变异密码几乎肯定是从中进化而来。因此，证据仍然很利于生命单一起源论。

怎样看待遗传密码是随机的，即现存密码无遗传优势的争论呢？实际上，考虑密码子在应对遗传突变的有效性时，我们发现它确实是很好的密码。首先，密码子中单碱基改变往往变成化学上类似的氨基酸。如亮氨酸、异亮氨酸和缬氨酸都有非常相似的疏水性侧链，它们的密码子也非常类似，只在第一个碱基上有所区别。所以，举一个特别好的例子，异亮氨酸密码子 AUA 的第一个碱基突变，可产生 UUA、CUA 或 GUA，前 2 个是亮氨酸密码子，后一个是缬氨酸密码子。因此，哪种突变都不会引起相应氨基酸的太大变化，因而减小了突变基因对蛋白质产物产生严重损伤的机会。

考虑其他两个因素时，我们的密码系统看上去更好：首先是**碱基转换**（transition）（嘌呤对嘌呤，或嘧啶对嘧啶的变化）比**碱基颠换**（transversion）（嘌呤变成嘧啶，或反之）普遍得多。第二是核糖体对密码子第一位和第三位碱基比第二位更容易误读。综合这些因素，对所有可能的 3 碱基密码，我们可以计算单碱基改变对编码氨基酸不产生影响或只有温和改变的概率（对出错的耐受性，译者注）。这样可以从总体上看出我们的天然密码与其他密码比较的优势。图 18.9 是这一数学分析的结果，表明我们的遗传密码确实是百万分之一，即在 100 万个其他可能的密码中，只有一个在减小突变的影响方面比我们的密码好。基于以上分析，遗传密码是偶然发生的，而不是进化精雕细磨的结果的可能性不大。

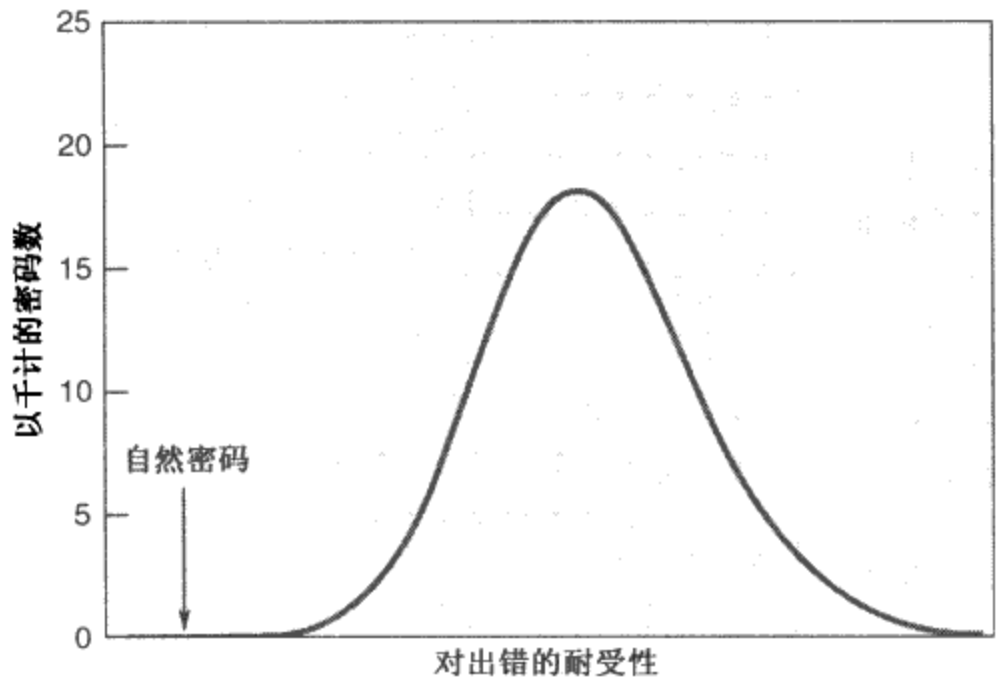


图 18.9　遗传密码对出错的耐受性。由 4 种碱基组成的所有可能三联遗传密码的出错耐受性对具有各耐受性值的密码数目（以千计）作图。我们的自然密码远离正态分布曲线的外侧，对出错的耐受性非常低。事实上，在百万中只有一个密码具有较低的耐受性。（*Source*：Adapted from Vogel, G. Tracking the history of the genetic code. *Science* 281（17 Jul 1998）329-331.）

小结　遗传密码并不是严格通用的，在某些真核生物的细胞核和线粒体基因组中、在至少一种细菌中，标准遗传密码的终止密码子可以编码氨基酸，如色氨酸和谷氨酰胺。在几种线粒体基因组以及至少一种酵母的核基因组中，密码子的含义从一种氨基酸变成了另一种氨基酸。这些变异的密码仍然与它们可能据此演变的标准密码子有着紧密的联系。现在还不清楚遗传密码的产生是冻结的偶然事件，还是进化的产物，但是其耐受突变的能力表明它是经受过进化的。

18.3　延伸机制

多肽链的延伸在不断重复的 3 个步骤中发生。我们先来概述这 3 个步骤，然后再了解细节以及实验证据。

延伸概述

图 18.10 简要地通过两轮延伸（在延伸的肽链上添加两个氨基酸）描述了*E. coli*中延伸的 3 个步骤。我们从 mRNA 和 fMet-tRNA$_f^{Met}$与核糖体的结合开始。核糖体上有 3 个氨酰-tRNA 结合位点。其中两个叫做 **P 位点**和 **A 位点**。在示意图中，P 位点在左侧，A 位点在右侧。fMet-tRNA$_f^{Met}$在 P 位点。去氨酰化-tRNA 的结合位点叫做 **E 位点**，为简洁起见在图中省略了。我们将在本章的后面介绍 E 位点。以下为图 18.10 所示延伸步骤的细节。

a. 为开始延伸，我们需要先加入另一个氨基酸，这个氨基酸到达并结合到 tRNA 上，该氨酰-tRNA 的

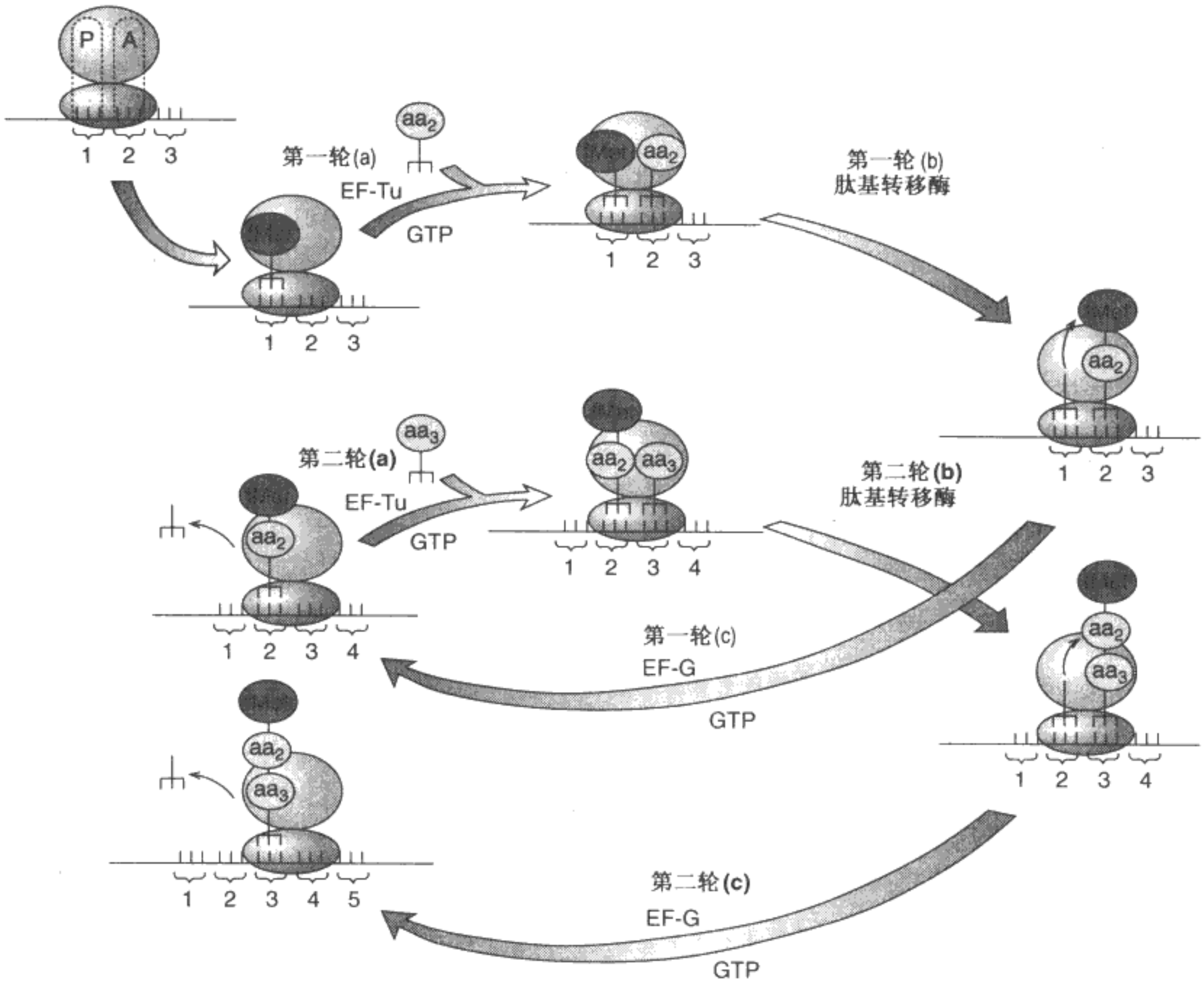


图 18.10　翻译延伸。首先要注意，这是蛋白质合成的高度示意图。例如，tRNA 用叉状结构表示，只显示分子的两个作用末端。左上方：与一条 mRNA 结合的核糖体显示 A 和 P 两个位点，以点线示出。第一轮：(a) 延伸因子 EF-Tu 将第二个氨酰-tRNA（黄色）带到核糖体的 A 位点，P 位点已经由 fMet-tRNA（紫红色）占据；(b) 肽酰转移酶在 fMet-tRNA 与第二个氨酰-tRNA 间形成肽键；(c) 在移位步骤，EF-G 将 mRNA 和 tRNA 向左移动一个密码子距离，使二肽酰-tRNA 进入 P 位，将去氨酰-tRNA 泵出，空出 A 位准备接受新的氨酰-tRNA。在第二轮，重复上述步骤添加一个新的氨基酸（绿色）。

性质由信使 RNA 的第二个密码子决定。第二个密码子位于 A 位点，否则该位点是空的，所以第二个氨酰-tRNA 结合到这个位点。这一结合过程需要延**伸因子**（elongation factor）EF-Tu 和 GTP。

b. 接着生成第一个肽键。**肽酰转移酶**（peptidyl transferase）（大核糖体亚基的整体部分）将 fMet 从 P 位点的 tRNA 上转移到 A 位点的氨酰-tRNA 上，在 A 位点形成一个叫做二肽的 2-氨基酸单位，与 A 位点的 tRNA 连接。A 位点的这个组装整体是个二肽酰-tRNA。留在 P 位点的是个去氨酰-tRNA（没有负载氨基酸的 tRNA）。

原核生物的肽酰转移酶步骤受一种重要的抗生素**氯霉素**（chloramphenicol）的抑制。该药物对多数真核生物的核糖体没有作用，使之能够选择性抑制高等生物中入侵的细菌。但是，真核生物线粒体有自己的基因组，氯霉素对其肽酰转移酶也能抑制。因此，氯霉素对细菌的选择性不是绝对的。

c. 接下来一步叫做**转位**（translocation）。mRNA带着 A 位上的肽酰-tRNA 向左移动一个密码子的长度，产生如下结果：①P 位点的去氨酰-tRNA（在肽酰转移酶催化肽键形成时失去氨基的 tRNA）通过 E 位点离开核糖体；②A 位的二肽酰-tRNA 及其相应的密码子移动到 P 位；③在右边等待的“箭在弦上”的密码子移到 A 位，准备接受新的氨酰-tRNA。转位需要 **EF-G** 延长因子外加 GTP。

然后重复这一过程，添加另一个氨基酸：(a) EF-

Tu 与 GTP 联合，使适配的氨酰-tRNA 与 A 位的新密码子配合；（b）肽酰转移酶使 P 位的二肽加入到 A 位的氨酰-tRNA 上，形成三肽酰-tRNA；（c）EF-G 将三肽酰-tRNA 及其 mRNA 密码子移位到 P 位点。

我们现在已经完成了两轮肽链延伸，从 P 位点的一个氨酰-tRNA（fMet-tRNA$_f^{Met}$）开始，使其延长了两个氨基酸成为三肽酰-tRNA。这个过程不断重复，直至核糖体阅读到 mRNA 的终止密码子，至此，多肽完成，该是链终止的时间了。

小结 肽链延伸以 3 个步骤重复发生：①EF-Tu 及 GTP 结合氨酰-tRNA 到核糖体的 A 位点；②肽酰转移酶催化 P 位的肽与 A 位新到来的氨酰-tRNA 间形成肽键，使肽链延长一个氨基酸，并将其移到 A 位点；③延伸因子 EF-G 以及 GTP 使延伸的肽酰-tRNA 及其 mRNA 密码子移位到 P 位点。

核糖体的 3 位点模型

为了便于理解，以上过程图中所示的核糖体都被简化了。简化之一是只给出两个 tRNA 结合位点（A 位和 P 位），虽然第三个位点（E 位）已有很好的证据。我们将从 A 位和 P 位的证据开始讨论，然后分析 E 位的证据。A 位和 P 位的存在最初基于抗生素**嘌呤霉素**（puromycin）实验（图 18.11）。该药物是偶联了一个腺嘌呤类似物的氨基酸，因此与氨酰-tRNA 尾部的氨酰腺嘌呤核苷类似，事实上，与结合在核糖体 A 位点的氨酰-tRNA 非常相似。它可以与 P 位的肽形成肽键，生成肽酰嘌呤霉素，到此，抗生素竞争 P 位点结束。肽酰嘌呤霉素与核糖体的结合不紧密，很快就被释放，导致翻译提前终止，这就是嘌呤霉素杀死细菌及其他细胞的原因。

嘌呤霉素与核糖体的 2-位点模型的联系在于：转位之前，A 位被肽酰-tRNA 占据，嘌呤霉素无法结合释放肽段；转位之后，肽酰-tRNA 移入 P 位，A 位空开，嘌呤霉素可以结合并释放肽段。因此，可以预计有两种状态的核糖体：嘌呤霉素反应的和嘌呤霉素非反应的。这两个状态要求在核糖体上至少有两个肽酰-tRNA 结合位点。

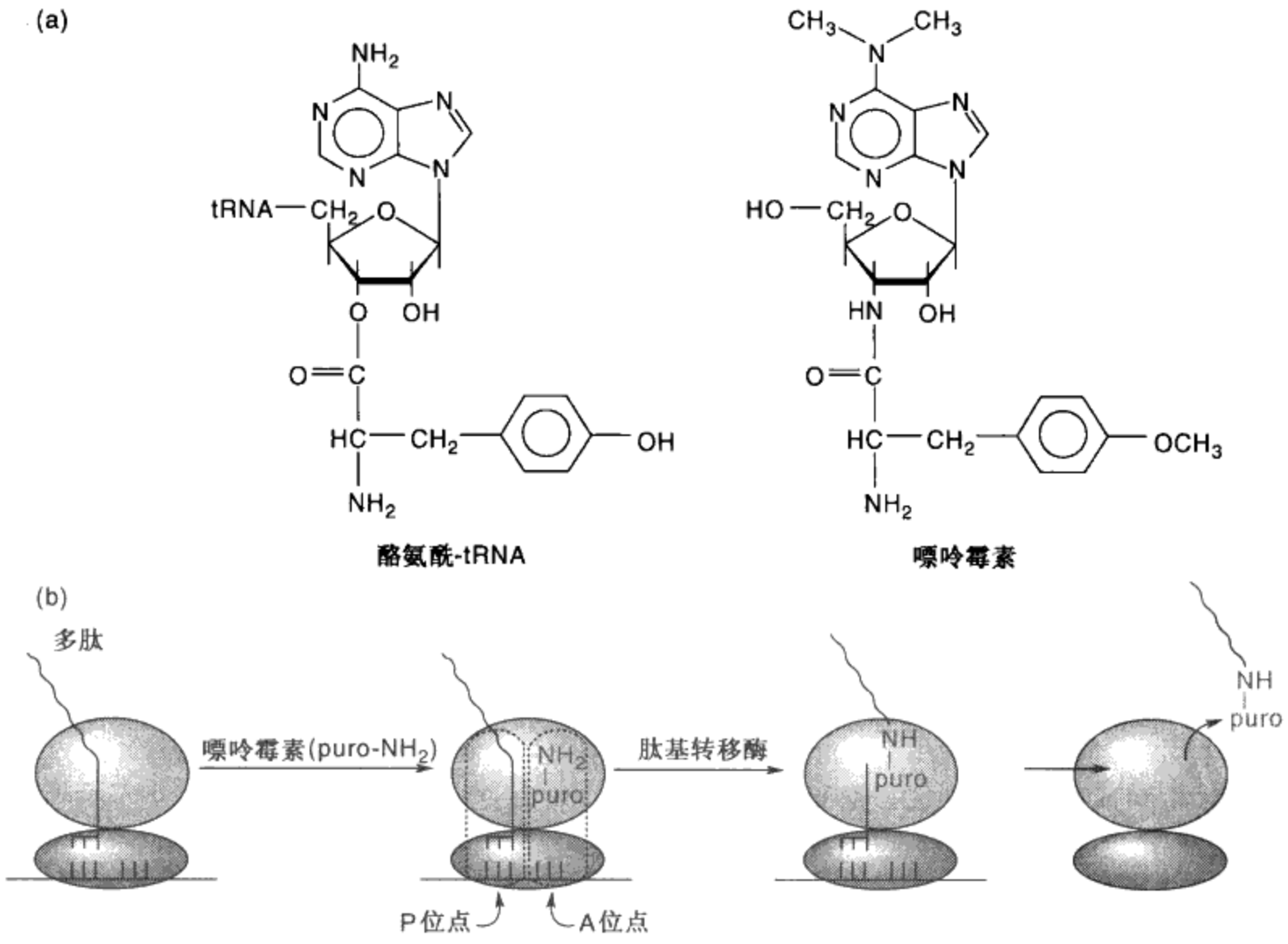


图 18.11 嘌呤霉素的结构与活性。（a）嘌呤霉素与酪氨酰-tRNA 结构的比较。注意，tRNA 的其他部分与氨酰-tRNA 的 5′碳原子结合，而在嘌呤霉素中只有一个羟基。嘌呤霉素与酪氨酰-tRNA 的差别用紫红色标出。（b）嘌呤霉素的作用模式。首先，嘌呤霉素（puro-NH_2）结合到核糖体空出的 A 位（A 位点必须空出给嘌呤霉素结合）。然后，肽酰转移酶将 P 位的肽连接到 A 位嘌呤霉素的氨基上。最后，肽酰-嘌呤霉素从核糖体上解离，提前终止翻译。

嘌呤霉素可用来显示氨基酰-tRNA是在A位还是P位。如果氨基酰-tRNA在P位，就能与嘌呤霉素形成肽键，并被释放；如果氨基酰-tRNA在A位，可阻止嘌呤霉素与核糖体结合，不能被释放。

这一相同过程可用来显示70S翻译起始复合物中fMet-tRNA是结合到P位点的。在第17章讨论（翻译）起始时，我们认为fMet-$tRNA_f^{Met}$结合到P位点，这是有道理的，因为这样才能将A位空出让后续氨基酸进入。1966年，M. S. Bretscher和Marcker用嘌呤霉素分析证明fMet-$tRNA_f^{Met}$确实是结合到P位点的。他们将［^{35}S］fMet-$tRNA_f^{Met}$与核糖体、三核苷酸AUG以及嘌呤霉素混合。如果AUG将fMet-$tRNA_f^{Met}$吸引到P位，那么标记的fMet将与嘌呤霉素反应，释放标记的fMet-嘌呤霉素。反之，如果fMet-$tRNA_f^{Met}$结合到A位，嘌呤霉素就不能结合，也就不会发生标记氨基酸的释放。图18.12显示，附着在$tRNA_f^{Met}$上的fMet确实被嘌呤霉素释放了，而附着$tRNA_m^{Met}$的Met却没有。因此，fMet-$tRNA_f^{Met}$结合到P位，而Met-$tRNA_m^{Met}$结合到A位。有人会认为可能是fMet而不是$tRNA_f^{Met}$造成实验结果的不同，为排除这种可能性，Bretscber和Marcker用Met-$tRNA_f^{Met}$做了相同的实验，发现其Met也被嘌呤霉素释放了（图18.2c），因此，证明是tRNA而不是甲硫氨酸上的甲酰基将氨酰-tRNA定位到P位点的。

1981年，Knud Nierhaus及同事提供了第三个核糖体位点E位点（E代表“Exit”）存在的证据。他们的实验策略是，分别用放射性的去苯丙氨酰-tRNA（$tRNA^{Phe}$缺Phe），或Phe-$tRNA^{Phe}$，或乙酰-Phe-$tRNA^{Phe}$与*E. coli*核糖体结合，并测定每个70S核糖体上结合的分子数。表18.2显示在poly（U）mRNA有或无的条件下实验的结果。一次只有1分子乙酰-Phe-$tRNA^{Phe}$与核糖体结合，结合部位可以是A位或P位。有2分子Phe-$tRNA^{Phe}$结合核糖体，

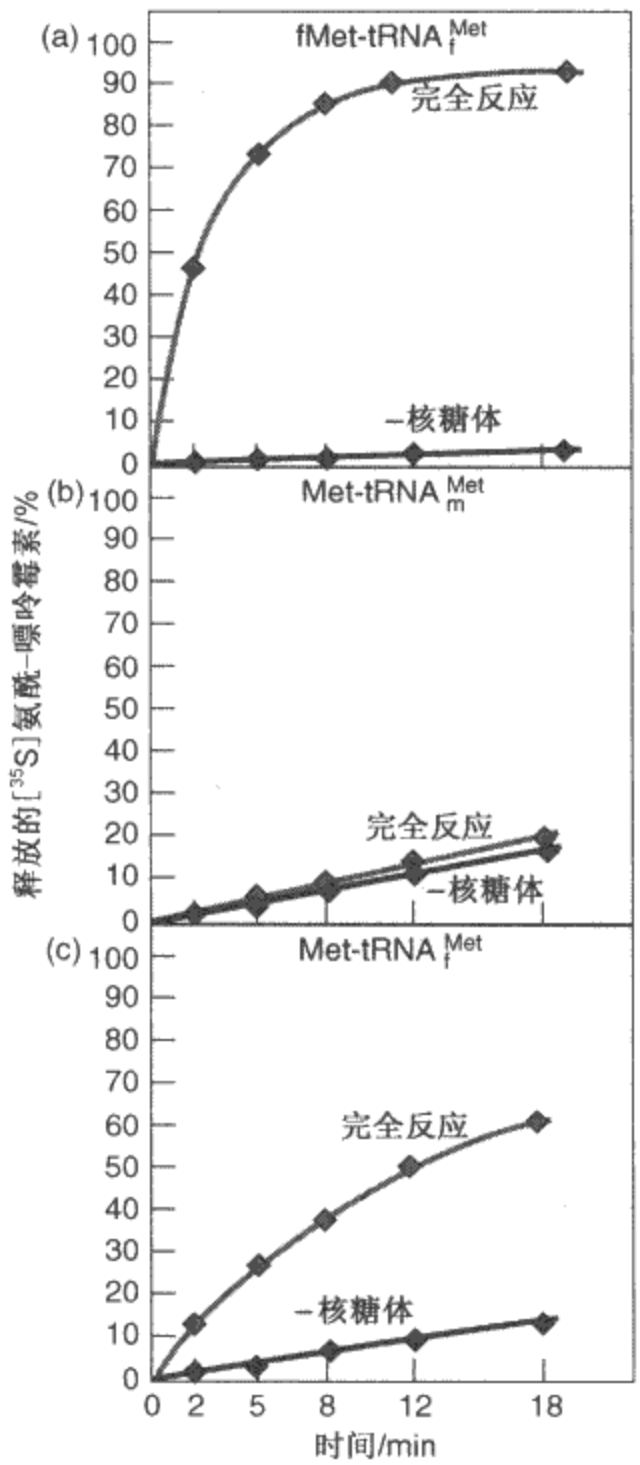


表18.2　tRNA和氨酰-tRNA对*E. coli*核糖体的结合

	tRNA	结合位点	
mRNA	种类	数量	位置
poly(U)	Acetyl-Phe-$tRNA^{Phe}$	1	P或A
poly(U)	Phe-$tRNA^{Phe}$	2	P和A
poly(U)	$tRNA^{Phe}$	3	P、E和A
无	$tRNA^{Phe}$	1	P
无	Phe-$tRNA^{Phe}$	0	—
无	Acetyl-Phe-$tRNA^{Phe}$	1	P

Source: Rheinberger, H.-J., H. Sternbach, and K. H. Nierhaus, Three tRNA binding sites on *Escherichia coli* ribosomes, *Proceedings of the National Academy of Sciences USA* 78(9): 5310-14, September 1981. Reprinted with permission.

图18.12　fMet-$tRNA_f^{Met}$占据核糖体P位点。Bretscher和Marcker利用嘌呤霉素-释放实验确定fMet-$tRNA_f^{Met}$在核糖体上的位置。分别用［^{35}S］fMet-$tRNA_f^{Met}$（a）、Met-$tRNA_m^{Met}$（b）、Met-$tRNA_f^{Met}$（c）与核糖体、AUG和嘌呤霉素混合，然后用高氯酸沉淀蛋白质和tRNA，检测标记的fMet-嘌呤霉素和Met-嘌呤霉素复合物的释放。从核糖体上释放的氨酰-嘌呤霉素酸是可溶的，而与核糖体结合的氨酰-tRNA则是酸不溶性的。完全反应包含所有成分，而对照反应则缺少其中一种成分，所缺少的成分都标注在每条曲线旁边。与$tRNA_f^{Met}$结合的fMet和Met位于P位点，能够被释放，与$tRNA_m^{Met}$结合的Met位于A位点，不能被嘌呤霉素释放。（*Source*: Adapted from Bretscher, M. S. and K. A. Marcker, Peptidyl-sRibonucleic acid and amino-acyl-sRibonucleic acid binding sites on ribosomes. *Nature* 211: 382-383, 1966.）

一个到 A 位，另一个到 P 位。3 分子的去苯丙氨酰-$tRNA^{Phe}$可以结合。如果在离开核糖体的路上有第三个位点与去氨酰-tRNA 结合，就可以容易地解释这些实验结果。因此，E 表示“出口”。在无poly（U）mRNA 时，只有一个 tRNA 以 $tRNA^{Phe}$，或乙酰-Phe-$tRNA^{Phe}$结合。Nierhaus 及同事估计结合位点是 P 位，随后的工作证实了他们的推测。

我们将在第 19 章详细讨论 E 位点，但这里也应注意，E 位点并不仅仅是去氨酰-tRNA 脱离核糖体过程中的一个位点，它在维持 mRNA 可读框中也起关键的作用。通常，可读框移位约为每 30 000 密码子发生一次，故 E 位点的作用很重要，因为可读框移位一般会产生无效蛋白质。但是某些 mRNA 的正确翻译却取决于可读框移位。

如*E. coli*的 *prfB* 基因，它编码 RF2，这是一种释放因子，我们将在本章后面介绍。为了使 *prfB* mRNA 能正确地翻译，必须向+1 位移位可读框。序列 CUUUGAC 正常是读为CUU UGA（Leu，终止）。+1 移位后，读为CUU*U*GAC（Leu，Asp），斜体的 U 被跳过了，下一个密码子是下画线的 GAC，编码天冬酰胺。

2004 年，Knud Nierhaus 及同事检测了体外 *prfB* mRNA 的翻译，发现 E 位点去氨酰 tRNA 的存在可阻止这种可读框移位。当去除 E 位的去氨酰 tRNA 后，可读框移位发生的频率非常高，于是他们得出结论：E 位点的去氨酰 tRNA 是维持正确可读框这一重要目的所正常需要的，当特殊 mRNA 的正确翻译需要可读框移位时，细胞必须将去氨酰 tRNA 从 E 位点移开。

小结 嘌呤霉素的化学结构与氨酰-tRNA 类似，能结合到 A 位点，与 P 位点的肽偶联并将其以肽酰-嘌呤霉素形式释放。此外，如果肽酰-tRNA在 A 位点，嘌呤霉素就不能与核糖体结合，肽段因而得不到释放。由此定义出核糖体上的两个位点：嘌呤霉素反应位点（P 位）和嘌呤霉素非反应位点（A 位）。在 70S 起始复合物上的 $fMet\text{-}tRNA_f^{Met}$具有嘌呤霉素反应特性，所以它位于 P 位点。结合实验鉴定出去氨酰-tRNA 结合的第三个位点（E 位点），预计此类 tRNA 在离开核糖体时与 E 位点结合。

延伸步骤 1：氨酰-tRNA 与核糖体 A 位点的结合

我们对延伸过程的详细了解始于 1965 年，当时 Yasutomi Nishizuka 和 Fritz Lipmann 用阴离子交换层析分离*E. coli*中肽键形成所需的两个蛋白质因子。他们将其中一个命名为 T（表示转移），因为它能将氨酰-tRNA 转移到核糖体上，将另一个叫做 G，因为其具有 GTP 酶活性（后面我们会看到 T 也有 GTP 酶活性）。然后，Jean Lucas-Lenard 和 Lipmann 证实 T 实际上由两种不同蛋白质组成，分别为 Tu（u 表示 unstable,）和 Ts（s 表示 stable）。这 3 个因子现在被分别称为 **EF-Tu**、**EF-Ts** 和 **EF-G**，它们参与延伸的第一步和第三步。我们首先讨论 EF-Tu 和 EF-Ts 的作用，因为它们参与延伸的第一个步骤。

1976 年，Joanne Ravel 发现未分离的 EF-T（EF-Tu 加 EF-Ts）具有 GTP 酶活性。另外，EF-T 结合氨酰-tRNA 到核糖体上需要 GTP。为证明这一现象，她用［^{14}C］Phe-$tRNA^{Phe}$与 EF-T 及浓度依次升高的 GTP 一起作用于核糖体，然后用硝酸纤维素薄膜过滤，结合了核糖体的标记 Phe-$tRNA^{Phe}$会留在滤膜上，而未结合的则被滤掉。图 18.13a 是实验结果。在没有延伸因子 EF-T 和 GTP 的情况下，Phe-$tRNA^{Phe}$对核糖体的非酶性背景结合较高，但是没有生理意义。忽略背景，我们可以看到 GTP 是 EF-T 依赖性 Phe-$tRNA^{Phe}$与核糖体结合所必需的。

当 Ravel 将 EF-T 和 EF-G 加入含有核糖体、在 poly（U）以及放射性标记的 Phe-$tRNA^{Phe}$存在溶液中时，她发现核糖体上生成了带标记的多聚苯丙氨酸。这种氨基酸多聚化作用所需的 GTP 浓度要比氨酰-tRNA 结合反应所需的浓度高。

在学习翻译的起始内容时，我们知道 IF-2 介导的 $fMet\text{-}tRNA_f^{Met}$与核糖体的结合也需要 GTP，但不需要 GTP 水解。那么对于 EF-T 及普通氨酰-tRNA 与核糖体的结合是否也是这样呢？1968 年，Anne-Lise Haenni 和 Lucas-Lenard 证明事实正是如此。他们用^{14}C和^{3}H分别标记 *N*-乙酰-Phe-tRNA 和 Phe-tRNA，然后将其与 EF-T 以及 GTP 或不能水解的类似物 GDPCP 混合。在非生理条件下，*N*-乙酰-Phe-tRNA 与核糖体 P 位点结合。研究人员用图 18.13 中介绍的过滤方法测定氨酰-tRNA 与核糖体的结合。他们还通过提取二肽产物和纸电泳鉴定了 P 位的 *N*-乙酰-Phe 与 A 位的 Phe-$tRNA^{Phe}$肽键的结合形式。表 18.3 显示，在 EF-T 及 GTP 或 GDPCP 帮助下，*N*-乙酰-Phe-$tRNA^{Phe}$可以结合到 P 位点，Phe-$tRNA^{Phe}$可以结合到 A 位点（事实上，*N*-乙酰-Phe-$tRNA^{Phe}$甚至不需要 EF-T 也能与 P 位结合）。因此，EF-T 促使氨酰-tRNA 与核糖体的结合不需要 GTP 的水解。明显不同的是，*N*-乙酰-Phe 与 *N*-乙酰-Phe-$tRNA^{Phe}$间肽键的形成需要 GTP 水解。这与翻译起始时的情况类似，IF-2 不需要 GTP 水解就可使 $fMet\text{-}tRNA_f^{Met}$结合到 P 位点，但随后的反应只有当 GTP 被降解后才能发生。

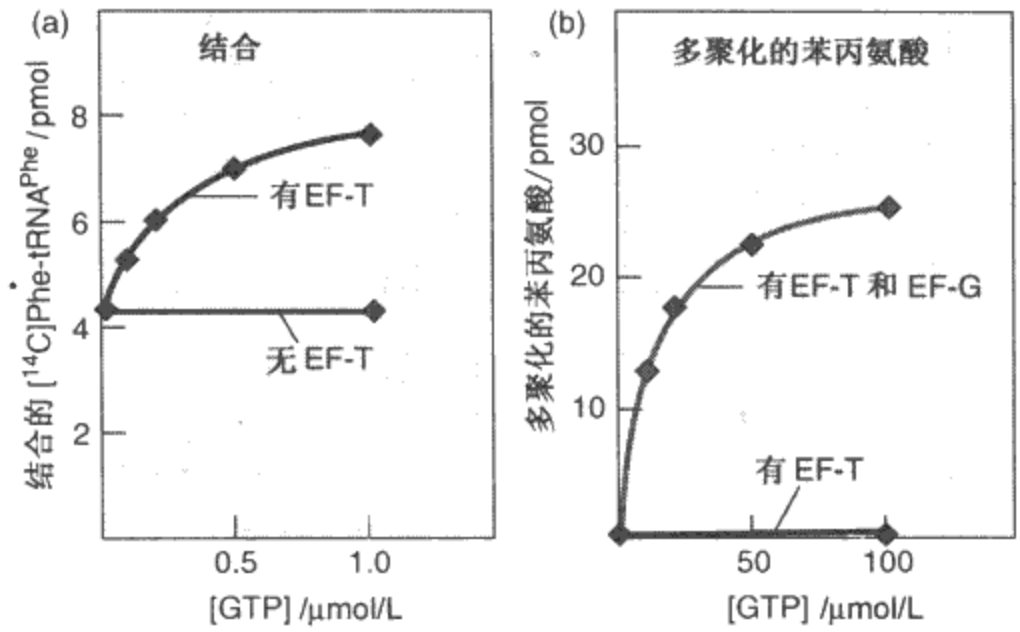


图 18.13　EF-T 和 GTP 对 Phe-tRNA^{Phe} 与核糖体结合以及多聚苯丙氨酸合成的影响。（a）Phe-tRNA^{Phe} 与核糖体的结合。在 EF-T 有或无的条件下，Ravel 在不同 GTP 浓度下将^{14} CPhe-tRNA^{Phe} 与核糖体混合。过滤法测定 Phe-tRNA^{Phe}-核糖体结合，以滤膜上标记的 Phe 与核糖体结合的量表示。EF-T 和 GTP 都不存在时，非酶性结合较高，但 EF-T 依赖性结合需要 GTP。（b）苯丙氨酸的多聚化。在 EF-G 有或无的条件下，将标记的 Phe-tRNA^{Phe} 在不同 GTP 浓度下与核糖体混合。酸沉淀测定多聚苯丙氨酸，步骤如下：用三氯乙酸（TCA）沉淀 poly（Phe），在 TCA 存在时，加热水解残留的 Phe-tRNA^{Phe}，过滤获得 poly（Phe）沉淀物。多聚化需要 EF-T、EF-G 和高浓度 GTP 的共同参与。（*Source*：Adapted from Ravel, J. M., Demonstration of a guanosine triphosphate-dependent enzymatic binding of aminoacyl-ribonucleic acid to *Escherichia coli* ribosomes. *Proceedings of the National Academy of Sciences USA* 57：1815，1967.）

他们还证实 EF-Tu 和 EF-Ts 都是 Phe-tRNA^{Phe} 与核糖体结合所需要的，所做的分析与表 18.3 相同，只是没有 GDPCP，并且 EF-Tu 和 EF-Ts 是分开的（EF-Tu 中有一点 EF-Ts 残留污染）并被分别加入。表 18.4 显示 EF-Tu 和 EF-Ts 都是 Phe-tRNA^{Phe} 与核糖体结合所需要的。只加 EF-Tu 时有少量结合是因为有 EF-Ts 污染。

表 18.3　GTP 和 GDPCP 对氨酰-tRNA 结合核糖体以及对结合和形成肽键的影响

添加成分	结合的 *N*-乙酰-Phe-tRNA^{Phe} (^{14}C)/pmol	形成的 *N*-乙酰-diPhe-tRNA(^{14}C 或^{3}H)/pmol	结合的 Phe-tRNA (^{3}H)/pmol
无	7.6	0.4	0.1
EF-T+GTP	3.0	4.5	2.8
EF-T+GDPCP	7.0	0.5	4.8

Source: Haenni, A. L. and J. Lucas-Lenard, Stepwise synthesis of a tripeptide, *Proceedings of the National Academy of Sciences*, *USA* 61:1365, 1968. Reprinted by permission.

表 18.4　[^{3}H] Phe-tRNA 结合到带有结合了 *N*-乙酰-[^{14}C] Phe-tRNA 的核糖体需要 EF-Ts 和 EF-Tu

添加成分	结合的[^{3}H]Phe-tRNA/pmol
无	2.8
EF-Ts+GTP	2.8
EF-Tu+GTP	5.2
EF-Ts+ EF-Tu+ GTP	11.6

Source: Naenni, A. L., and J. Lucas-Lenard, Stepwise synthesis of a tripeptide, *Proceedings of the National Academy of Sciences USA* 61:1365, 1968. Reprinted with permission.

图 18.14 是 EF-Tu 与 EF-Ts 共同作用引起氨酰-tRNA 与核糖体结合的详细机制模型。首先，EF-Tu 与 GTP 构成一个二元（2 部分）复合物。然后，氨酰-tRNA 加入这个复合物，形成由 EF-Tu、GTP、氨酰-tRNA 构成的三元（3 部分）复合物。该三元复合物将其氨酰-tRNA 送到核糖体的 A 位点，EF-Tu 和 GTP 保持结合在核糖体上。随后，GTP 被水解，EF-Tu-GDP 复合物从核糖体上解离。最后，EF-Ts 使复合物上的 GDP 与 GTP 交换，产生 EF-Tu-GTP 复合物。

以上过程的证据是什么呢？Herbert Weissbach 及同事在 1967 年发现，EF-T 制备物和 GTP 形成的复合物可以被硝酸纤维素滤膜截留。他们标记了 GTP，将其与 EF-T 混合，发现标记的核苷酸结合到滤膜上了，这表明 GTP 已经结合到 EF-T 制备物的某个蛋白质上，可能是 EF-T 本身，形成了复合物。Julian Gordon 随后发现，将氨酰-tRNA 加入 EF-Tu-GTP 复合物可使其从滤膜上释放。一种解释是氨酰-tRNA 结合 EF-Tu-GTP 复合物形成四元复合物从而不再与滤膜结合。有两条证据支持这一假设。第一，去氨酰-tRNA（缺失氨基酸的 tRNA）不能从滤膜上释放 EF-T-GTP 复合物（图 18.15），这一条满足假设是因为去氨酰-tRNA 不能进入将氨酰-tRNA 送到核糖体上的复合物中。此外，其他研究者发现 fMet-tRNA$_{f}^{Met}$、Met-tRNA$_{f}^{Met}$、*N*-乙酰-Phe-tRNA^{Phe} 都与去氨酰-tRNA 一样不能从滤膜上释放 EF-T-GTP 复合物。因此，只有注定要结合到 A 位点的氨酰-tRNA 才可以与 EF-T 和 GTP 结合形成四

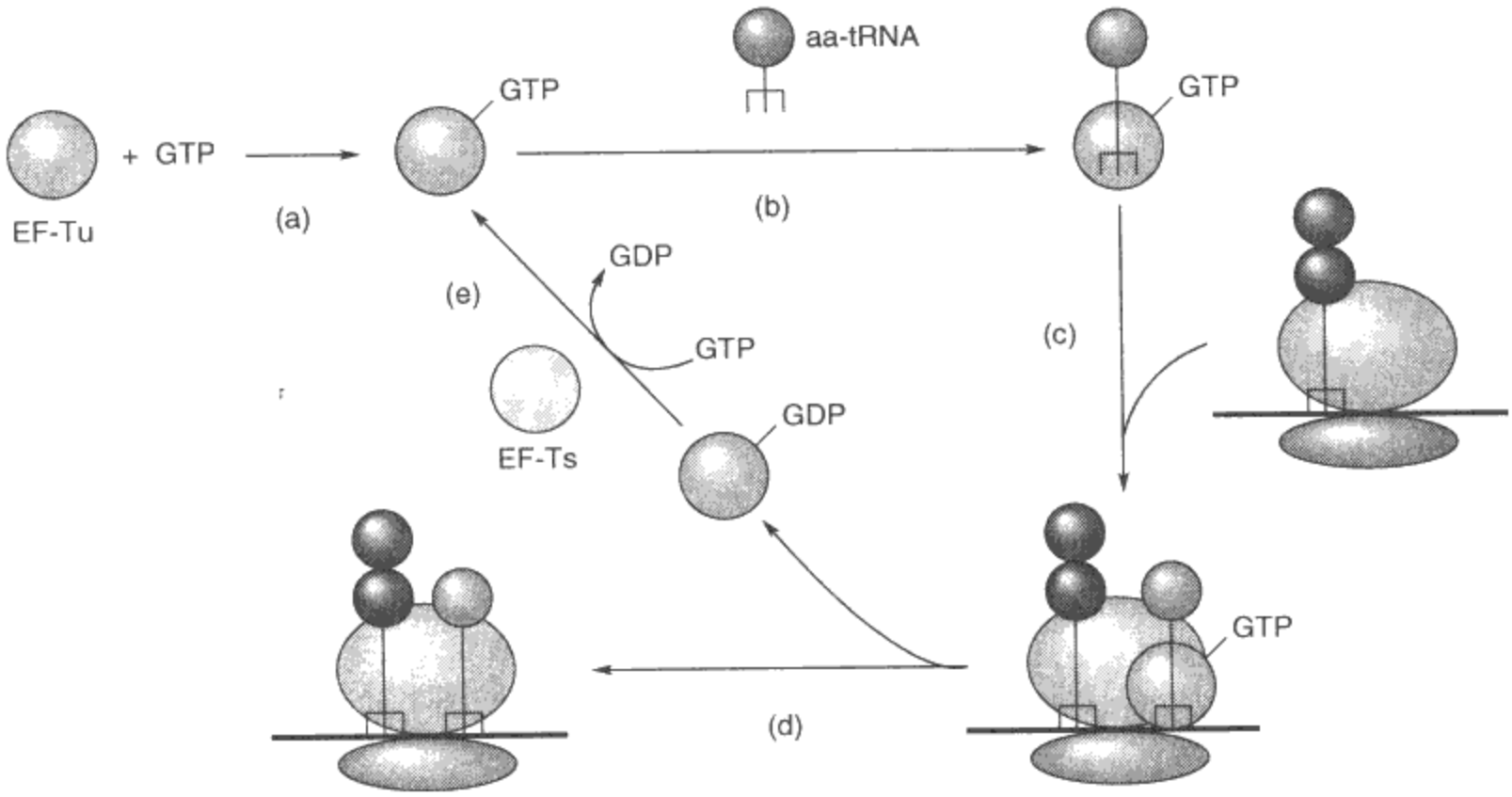


图 18.14　氨酰-tRNA 与核糖体 A 位点的结合模型。（a）EF-Tu 与 GTP 结合形成二元复合物。（b）二元复合物又与一个氨酰-tRNA 结合，形成三元复合物。（c）三元复合物与 P 位是肽酰-tRNA、A 位空出的核糖体结合。（d）GTP 被水解，形成的 EF-Tu-GDP 复合物从核糖体上解离，在 A 位留下新的氨酰-tRNA。（e）EF-Ts 用 GTP 交换 EF-Tu 的 GDP 再生 EF-Tu-GTP 复合物。（aa-tRNA 二氨酰-tRNA）

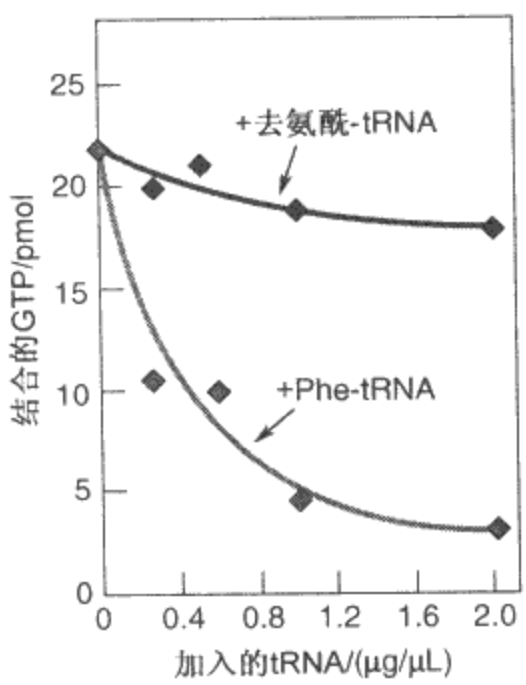


图 18.15　氨酰-tRNA 和去氨酰-tRNA 对 EF-Tu-GTP 复合物的影响。Gordon 分别将 Phe-tRNA^{Phe} 和去氨酰-tRNA 与含有［^{3}H］GTP 的 EF-Tu-GTP 复合物混合，将混合物通过硝酸纤维素滤膜。用结合在滤膜上的复合物中 GTP 的量对加入的 tRNA 的量作图。Phe-tRNA^{Phe} 引起与滤膜结合复合物的大幅度下降，而去氨酰-tRNA 几乎无影响。（*Source*：Adapted from Gordon，J. A stepwise reaction yielding a complex between a supernatant fraction from *E. coli*，guanosine 5′-triphosphate，and aminoacyl-sRNA，*Proceedings of the National Academy of Sciences USA* 59：180，1968.）

元复合物。支持存在四元复合物的第二个证据是加入氨酰-tRNA 时，EF-T 的活性可以穿过滤膜进入滤液（图 18.16）。这表明 EF-T 是二元复合物的一部分，而且从标记实验我们已经知道 GTP 是该复合物的一部分。由于 EF-T 自身或只与 GTP 结合就可与滤膜结合，因此可以认为它之所以流过滤膜是因为存在于与 GTP 和氨酰-tRNA 的复合物中。

Ravel 及合作者用以下实验给我们提供了形成四元复合物的更多证据。他们用^{3}H 和^{32}P 标记 GTP，用^{14}C 标记 Phe-tRNA^{Phe}，与 EF-T 混合后，在 SephadexG100（第 5 章）进行凝胶过滤。这种凝胶过滤树脂排除大分子蛋白质，如 EF-T，所以它们在外水体积（void volume）组分中迅速流过层析柱。相反，相对小的物质，如 GTP，甚至 Phe-tRNA^{Phe} 进入树脂的孔中，因而被阻滞，在外水体积之后较晚出现。实际上，分子越小，洗脱出来所需的时间越长。图 18.17 是凝胶过滤实验的结果，标记物 GTP 和 Phe-tRNA^{Phe} 组分的出现相对于它们通常的位置较晚，表明这些组分是游离的 GTP 和 Phe-tRNA^{Phe}，虽然本实验中几乎没有观察到游离的 GTP。然而，实验中绝大多数的两种物质的组分出现得很早（在组分数 20 前后），表明它们一定与某种较大的物质形成了复合物。实验中主要的大物质是 EF-T，我们已经讨论过的实验提示该复合物中有 EF-T，所以可以推测形成了包括 Phe-tRNA^{Phe}、GTP、EF-T

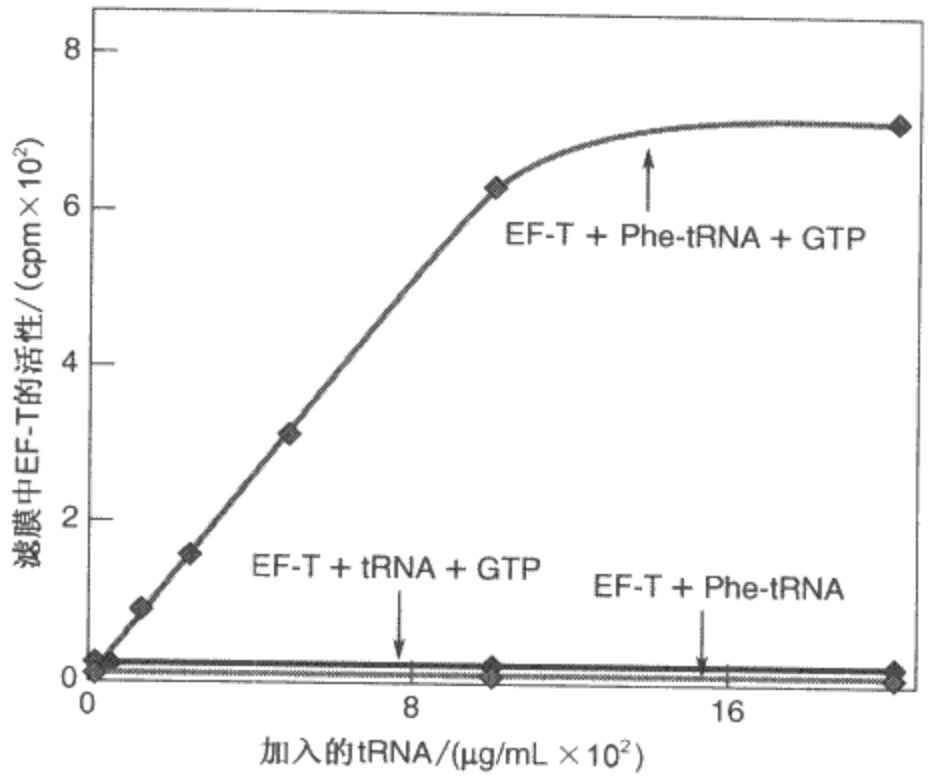


图 18.16　加入氨酰-tRNA 后硝酸纤维素滤液中 EF-T 的变化。Gordon 将下列成分混合并通过硝酸纤维素滤膜，利用苯丙氨酸多聚化分析检测滤液中 EF-T 的活性。红色：EF-T＋Phe-tRNA^{Phe}＋GTP；蓝色：EF-T＋去氨酰-tRNA＋GTP；绿色：EF-T＋Phe-tRNA。只有当 EF-T、Phe-tRNA 和 GTP 都存在时，EF-T 复合物才能穿过滤膜。（*Source*：Adapted from Gordon, J., A stepwise reaction yielding a complex between a supernatant fraction from *E. coli*, guanosine 5′-triphosphate, and aminoacyl-sRNA, *Proceedings of the National Academy of Sciences USA* 59：182，1968.）

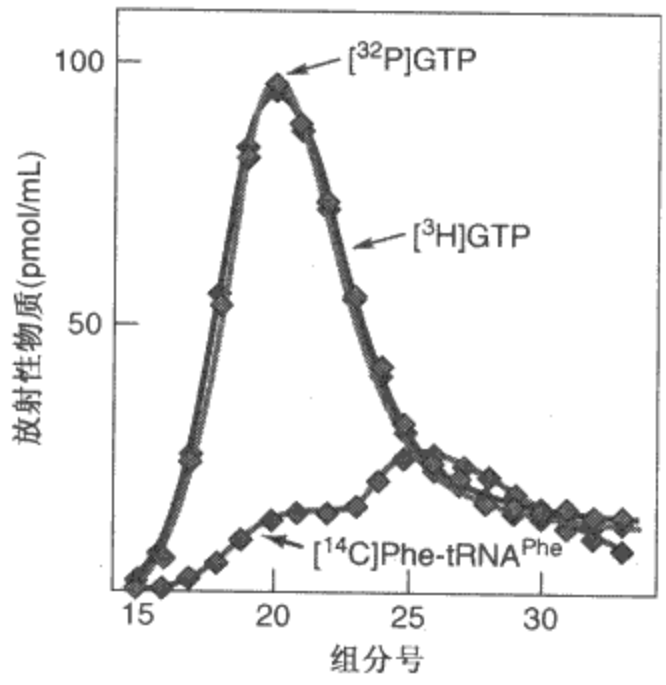


图 18.17　EF-T、氨酰-tRNA、GTP 三元复合物的形成。Ravel 及同事将［^{14}C］-Phe-tRNA^{Phe} 与 GTP（用^{3}H 标记鸟嘌呤、^{32}P 标记 γ-磷酸基团）和 EF-T 混合。混合物通过 SephadexG100 凝胶过滤层析柱，使大分子的 EF-T 与相对小分子的 GTP 和 Phe-tRNA^{Phe}分离。通过对每一组分中的 3 种放射性同位素的分析，测定 GTP 和 Phe-tRNA^{Phe}。在大分子组分中（组分数 20 左右），发现二者至少有部分存在，因此，它们确实与 EF-F 形成了复合物。（*Source*：Adapted from Ravel, J. M., R. L. Shorey, and W. Shive, The composition of the active intermediate in the transfer of aminoacyl-RNA to ribosomes. *Biochemical and Biophysical Research Communications* 32：12，1968.）

的三元复合物。

至此，在这些实验中我们还没有鉴别出 EF-Ts 和 EF-Tu。Herbert Weissbach 及其合作者通过将这两种蛋白质分离开，并分别进行检测完成了这一工作。他们发现 EF-Tu 是二元复合物中结合 GTP 的因子。那么 EF-Ts 的作用是什么呢？研究者证明，该因子在 EF-Tu-GDP 复合物向 EF-Tu-GTP 复合物转换中起关键作用。但是，当 EF-Ts 与事先形成的 EF-Tu-GTP 复合物或与 EF-Tu 单独在一起时，几乎没有作用（图 18.18）。因此，EF-Ts 可能并不直接与 EF-Tu 和 GTP 形成复合物，相反，是通过交换鸟嘌呤核苷酸使 EF-Tu-GDP 转换成 EF-Tu-GTP 的。

EF-Ts 如何执行它的交换任务呢？David Miller 和 Weissbach 发现，EF-Ts 通过形成 EF-Ts-EF-Tu 复合物可以从 EF-Tu-GDP 中置换 GDP（图 18.19）。这种置换作用怎样发生呢？Reuben Leberman 及同事对 EF-Ts-EF-Tu 复合物的 X 晶体学研究表明，EF-Ts 结合 EF-Tu-GDP 的主要结果之一是对 EF-Tu 的 Mg^{2+} 结合中心的破坏，EF-Tu 与 Mg^{2+} 间结合减弱导致了 GDP 的解离，从而给 GTP 结合 EF-Tu 开通了道路。

为什么需要 EF-Tu 护送氨酰-tRNA 到核糖体上呢？连接氨基酸与其同类 tRNA 的酯键很容易断裂，将氨酰-tRNA 隐藏在 EF-Tu 蛋白中可以保护这个易分解的化合物不被水解。但是细胞中氨酰-tRNA 的浓度很高，有足够的 EF-Tu 供应吗？是的，EF-Tu 是细胞中最丰富的蛋白质之一。例如，在*E. coli*细胞中 EF-Tu 占总蛋白的 5%，出现这种高丰度的原因可能是 EF-Tu 起重要的保护作用。

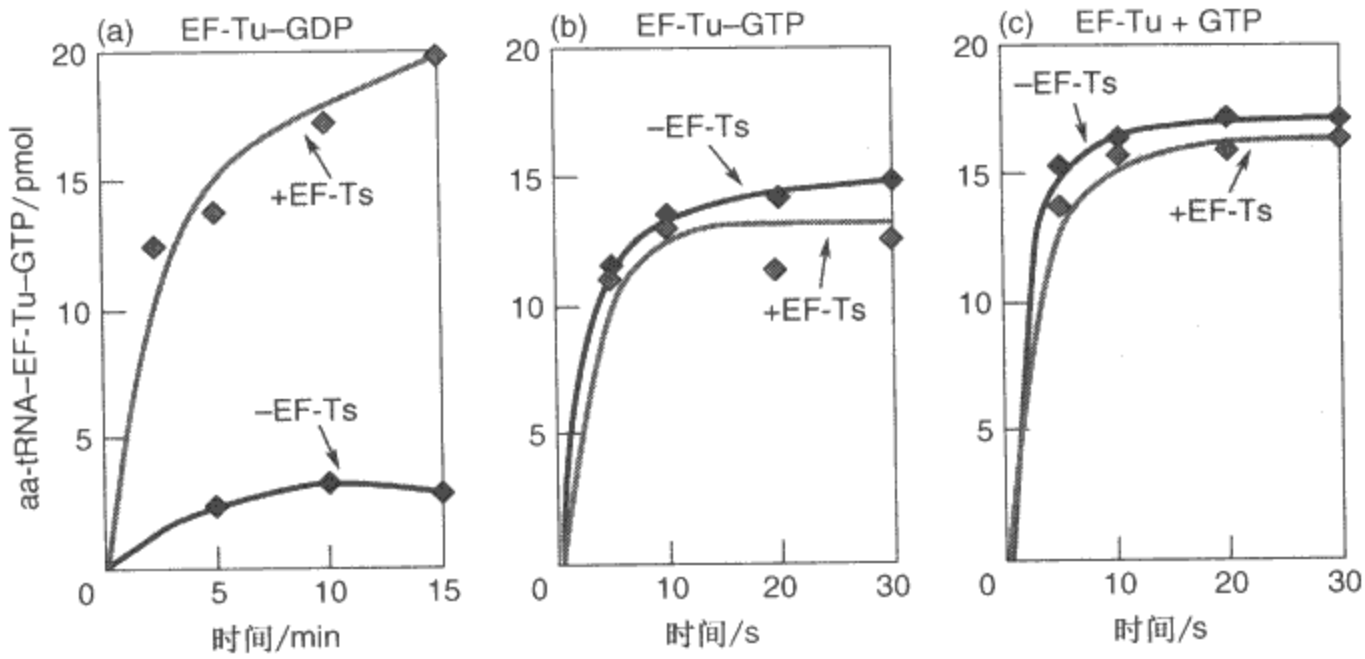


图 18.18　EF-Ts 对三元复合物形成的作用。Weissbach 及同事在有（红色）和无 EF-Ts（蓝色）的条件下，将 [^{14}C] Phe- tRNA^{Phe}、[^{3}H] GTP、EF-Tu 制备物混合，形成三元复合物。检测硝酸纤维素滤膜上放射性的减少来测定三元复合物的形成。只有当 EF-Tu-GDP 为底物时，EF-Ts 才能激发复合物的形成（图 a）。EF-Tu-GTP（图 b）或 EF-Tu＋GTP（图 c），无须 EF-Ts，就能自发形成三元复合物。（aa-tRNA＝氨酰-tRNA）（*Source*：Adapted from Weissbach，H.，D. L. Miller，and J. Hachmann，Studies on the role of factor Ts in polypeptide synthesis. *Archives of Biochemistry and Biophysics*，137：267，1970.）

小结　由 EF-Tu、氨酰-tRNA 和 GTP 形成的三元复合物，将氨酰-tRNA 转移到核糖体的 A 位点，此过程不需要 GTP 水解。随后，EF-Tu 以其核糖体依赖性 GTP 酶活性水解 GTP，形成的 EF-Tu-GDP 复合物从核糖体解离。EF-Ts 通过以 GTP 交换与 EF-Tu 结合的 GDP，使 EF-Tu-GTP 复合物再生。氨酰-tRNA 的加入又重新构成三元复合物进入新一轮肽链延伸。

校正　在第 19 章我们将看到蛋白质合成的准确性部分是来自 tRNA 对正确氨基酸的负载，部分是来自延伸步骤 1 中核糖体 A 位点密码子与氨酰-tRNA 的正确识别。但是，如果在这个初始识别步骤发生错误，仍有机会在错误氨酰-tRNA 将其氨基酸送到延伸链之前来纠正。这个过程称为**校正**（proofreading）。

Robert Thompson 用图 18.20 显示的过程总结了这些概念。k 值是每个反应的速度常数。这些速度常数可发生变化，取决于核糖体结合的三元复合物是否带有正确或不正确的氨酰-tRNA。从正确到不正确变化最大的是 k_{-1} 和 k_4。前者表示三元复合物与核糖体结合之后的解离速度。如果在三元复合物中氨酰-tRNA 是正确的，就不应该发生解离，但实际上解离以很低的速度发生着。如果复合物中氨酰-tRNA 不正确，就会以很高的速度发生解离，因为密码子与反密码子之间碱基配对不好，不能像完好配对那样紧密结合不正确的三元复合物。通常不正确三元复合物的解离发生在 GTP 水解之前，因此，其氨酰-tRNA 通常都不会留在核糖体上。

第二个速度常数 k_4 表示氨酰-tRNA（来自三元复合物）从核糖体上解离的速度。同样，如果氨酰-tRNA 不正确，那么解离发生的速度比正确情况下快得多，因为密码子与反密码子之间不完全配对。这种解离速度通常非常快，使不正确氨酰-tRNA 在其氨基酸有机会加入新生的肽链之前就与核糖体解离了。

分析翻译的准确性得出一个一般性原则，即高准确性与高翻译速度是不相容的。事实上，准确性与速度是负相关的：翻译速度越快，准确性越差。这是因为核糖体必须有足够的时间让不正确三元复合物和氨酰-tRNA 在不正确氨基酸不可逆转地加入延伸的肽链之前离开。如果翻译速度加快，就会有更多不正确的氨基酸加入。相反，如果翻译速度慢，准确性就会更高，但是蛋白质也许就不能足够快地合成以维持生命。因此，翻译的速度和准确性之间存在着精细的平衡。

在这个平衡中最重要的因素之一是 EF-Tu 对 GTP 的水解速度。如果水解速度快，则第一个校正的时间就短了：EF-Tu 很快将 GTP 水解成 GDP，没有给携带错误氨酰-tRNA 的三元复合物充足的时间与核糖体解离。另一方面，如果水解速度较慢，校正会有充裕的时间，但翻译速度会太慢。那么合适的速度是多少呢？在 *E. coli* 中，三元复合物的结合与 GTP 水解之间的平均时间大约是几毫秒。但是

EF-Tu-GDP 从核糖体解离需要更多个毫秒。校正就发生在这两个时间段内，缩短这些时间都会严重影响翻译的准确性。

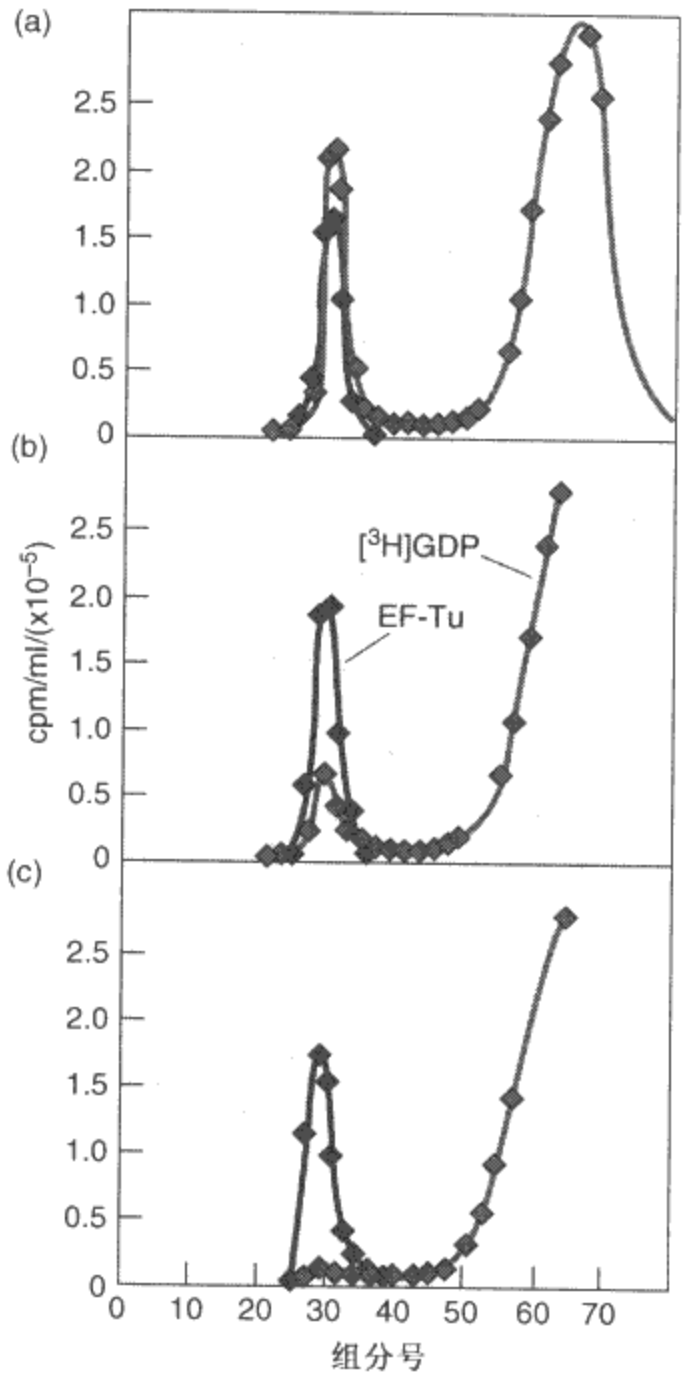


图 18.19　EF-Ts 对 EF-Tu-GDP 复合物中 GDP 的置换。Miller 和 Weissbach 将 EF-Tu-［^{3}H］GDP 复合物与 3 种不同浓度的 EF-Ts 混合，将混合物通过 SephadexG-25 凝胶过滤柱，测定复合物中 GDP 的含量。3 个实验组中 EF-Ts 的浓度分别为：（a）500 单位；（b）14 000 单位；（c）25 000 单位。红色，^{3}H-GDP；蓝色，EF-Tu。（*Source*：Adapted from Miller，D. L. and H. Weissbach，Interactions between the elongation factors：The displacement of GDP from the Tu-GDP complex by factor Ts. *Biochemical and Biophysical Research Communications* 38：1019，1970.）

细胞能忍受多大的翻译错误率呢？例如，错误率为 1%时会怎么样？99%的正确率听起来相当好了，但这是在你意识到大多数多肽链都远远超过 100 个氨基酸之前。多肽链的平均长度为 300 个氨基酸，有些甚至超过 1000 个氨基酸。假设每个氨基酸的错误率为 ε，多肽链长度为 n，则翻译一条无错（error-free）多肽链的概率 p 用下列公式计算：

$$p = (1-\varepsilon)^n$$

例如，以错误率为 1%，那么一条均长多肽链一次翻译不出错的概率只有 5%，而一条 1000 个氨基酸的多肽，不出错的概率几乎为零。如果错误率为 0.1%，那么正确翻译均长多肽的概率为 74%，而 1000 个氨基酸的多肽不出错的概率只有 37%，对于大分子多肽来说，这显然是一个问题。如果错误率为 0.01%，情况又会怎样？此时，正确翻译均长多肽的概率为 97%，而 1000 个氨基酸的多肽约为 91%。这是一个可接受的出错率。事实上，至少在*E. coli*中，观察到的加入每个氨基酸的错误率接近 0.01%。

一种重要的抗生素**链霉素**（streptomycin）会干扰校正，使核糖体产生很多错误。例如，正常核糖体翻译人工合成的 poly（U）只会单一地产生多聚苯丙氨酸。但是，链霉素可极大地刺激 Ile 和少量的 Ser 及 Leu 的加入。

某些自然条件使我们可以观察翻译速度比正常更快或更慢时的影响。例如，核糖体蛋白 *ram* 突变，或 EF-Tu 内部如 *tufAr* 突变都会使肽键形成速率 k_3 值增大 1 倍。在这些突变中，由于错误氨基酰-tRNA 不能及时从核糖体上解离，翻译的准确性下降。

相反，在链霉素抗性突变体如 *strA* 中，k_3 只有正常值的一半，因此，肽键形成的速度也只有正常速度的一半，使错误氨酰-tRNA 有充裕的时间离开核糖体，所以翻译格外准确。

小结　蛋白质合成机器在延伸过程中，通过一个两步过程来保证准确性。首先，在 GTP 水解之前，清除携带错误氨酰-tRNA 的三元复合物。如果这个监控失败，仍能在错误氨基酸加入延伸链之前清除错误的氨酰-tRNA。根据 Thompson 的假说，这两个校正过程依赖于不正确密码子与反密码子之间的碱基弱的配对作用，从而确保解离作用比 GTP 水解或肽键的形成更快发生。翻译的速度和准确性之间的平衡是精密的。如果肽键形成过快，不正确氨酰-tRNA 不能及时离开核糖体，它们的氨基酸就会被整合到蛋白质中，如果肽键形成速度过慢，蛋白质的合成又不能满足生物体生长的需要。每加入一个氨基酸约有 0.01%的出错率，可以实现翻译速度与准确性之间的良好平衡。

延伸步骤 2：肽键的形成

在起始因子和延伸因子 EF-Tu 完成它们的工作之后，核糖体在 P 位点就拥有了 fMet-tRNA$_f^{Met}$，在

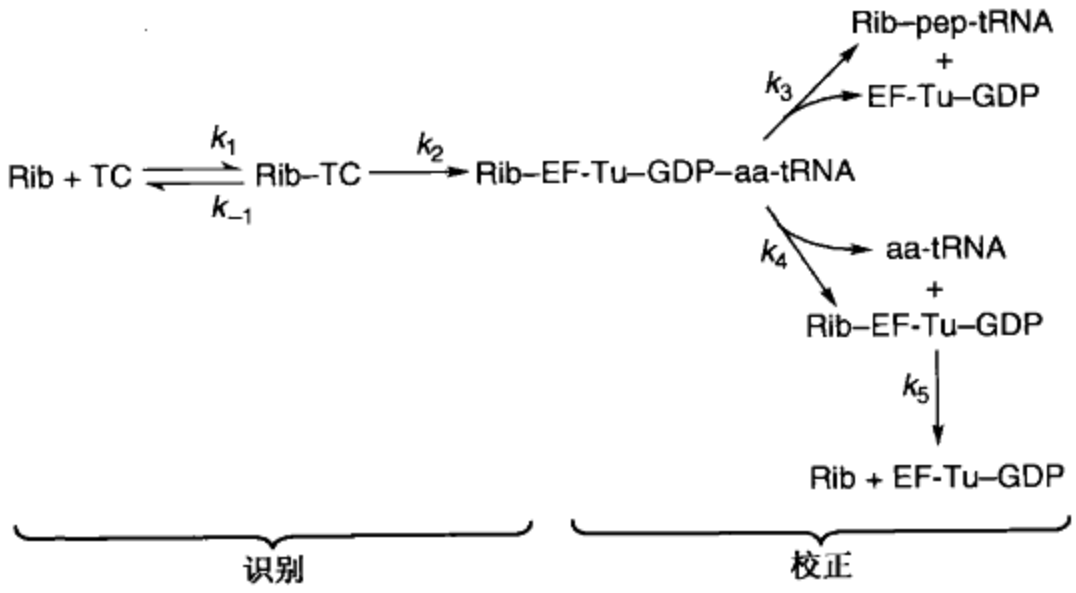


图 18.20　氨基酸掺入蛋白质的可能机制。每个速度常数表示箭头所指反应的速度。例如，k_1表示三元复合物（TC）结合到核糖体（Rib）上的速度；k_{-1}表示逆反应速度。（*Source*：Adapted from Thompson，R. C.，EFTu provides an internal kinetic standard for translational accuracy. *Trends in Biochemical Sciences* 13：91-93，1988.）

A 位点拥有氨酰-tRNA，现在该是形成肽键的时候了。你可能会期望一组新的延伸因子参与这个事件，但实际上没有。相反，核糖体自身就具有**肽基转移酶**（peptidyl transferase）活性，可形成肽键而不需要可溶性因子。

经典的肽基转移酶分析是由 Robert Traut 和 Robert Monro 发明的，用的是结合到核糖体 P 位的标记氨酰-tRNA 或肽酰-tRNA 和嘌呤霉素。标记的氨酰-嘌呤霉素或肽酰-嘌呤霉素的释放，依赖于 P 位点氨基酸或肽与 A 位点嘌呤霉素间肽键的形成，如图 18.21a 所示。Traut 和 Monro 还发现这个反应系统经过一定修饰后，可以显示 50S 核糖体亚基不需要 30S 亚基或其他可溶性因子的帮助，就可以执行肽基转移酶反应（图 18.21b）。首先，让核糖体以 poly（U）为 mRNA 进行 poly（Phe）合成，使 poly（Phe）-tRNA 进入 P 位点，然后通过与低浓度 Mg^{2+} 缓冲液孵育，再超速离心去除 30S 亚基。再用盐溶液洗去残留的起始因子和延伸因子，留下 50S 亚基结合 poly（Phe）-tRNA。通常，这种引发的 50S 亚基与嘌呤霉素不起反应，但研究者发现，加入 33%的甲醇（乙醇也有效）可以引起嘌呤霉素的反应。在两种分析中，首先必须区分被释放的肽酰-tRNA 和仍然与核糖体结合的肽酰-tRNA。Traut 和 Monro 最初采用蔗糖梯度离心进行分离（图 18.22）。

后来，又开发出更方便的过滤结合分析法。图 18.22a 是不含嘌呤霉素的阴性对照，如所预期的，poly（Phe）仍然结合在 50S 亚基上。图 18.22b 是阳性对照，加入尿素和 RNase 使 poly（Phe）从 50S 亚基上释放出来了。图 18.22c 和 18.22d 是嘌呤霉素分别加或不加 GTP 的实验结果。嘌呤霉素将 poly（Phe）释放出来了，表明肽基转移酶有效。即使没有 GTP，这个反应也发生了，正如肽基转移酶反应所应该的那样。

嘌呤霉素与 50S 亚基的反应似乎证明 50S 亚基含有肽基转移酶活性，那么非生理性条件（33%甲醇、嘌呤霉素）能破坏这一性质吗？令人鼓舞的是，肽与嘌呤霉素的反应好像遵守正常的肽合成的机制，而且 M. A. Gottesmann 用 poly（A）替换 poly（U），因而 poly（Lys）替换 poly（Phe），赖氨酰-tRNA 替换嘌呤霉素发现了相同的反应，证明嘌呤霉素反应是研究肽键形成的有效模型。此外，这些反应都受氯霉素和其他抑制正常肽基转移酶反应的抗生素的阻断，显示模型反应使用与正常反应相同的途径。

在以后的几十年中，一直不清楚 50S 亚基的哪个部位具有肽基转移酶活性。但是，当 Thomas Cech 及同事在 20 世纪 80 年代初发现某些 RNA 具有催化活性后，分子生物学家就开始怀疑 23S rRNA 可能催化了肽基转移酶的反应。1992 年，Harry Noller 及同事提供了这一推测的证据。在进行肽基转移酶分析时，他们使用了一种经修饰的嘌呤霉素反应叫做**片段反应**（fragment reaction）。反应由 Monro 在 20 世纪 60 年代开创，利用了在 P 位点的标记 fMet-$tRNA_f^{Met}$ 片段和 A 位点的嘌呤霉素。片段可以是 CCA-fMet 或 CAACCA-fMet，二者都与整个 fMet-$tRNA_f^{Met}$ 非常类似，因此都可以和 P 位点结合。随后，标记的 fMet 可与嘌呤霉素反应释放标记的 fMet-嘌呤霉素。

Noller 及合作者面临的任务是除去 50S 亚基上的所有蛋白质分子，只留下 rRNA，证明这个 rRNA 分子能够催化片段反应。为了去除 50S 亚基上的蛋白质，他们使用了 3 种较为强烈的、可以使蛋白质降解和变性的试剂：苯酚、SDS 和蛋白酶 K（PK）。图 18.23 中，1～4 泳道显示*E. coli* 50S 亚基经 SDS 和蛋白酶 K 处理可保留肽基转移酶活性，但苯酚处理的提取物则没有。这种能耐受 SDS 和蛋白酶 K 的能力很了不起，但是为什么苯酚提取物比其他两种物质对肽基转移酶的破坏更大？

Noller 及同事推测苯酚可能破坏了对肽基转移酶活性非常重要的 RNA 的高级结构。如果是这样，

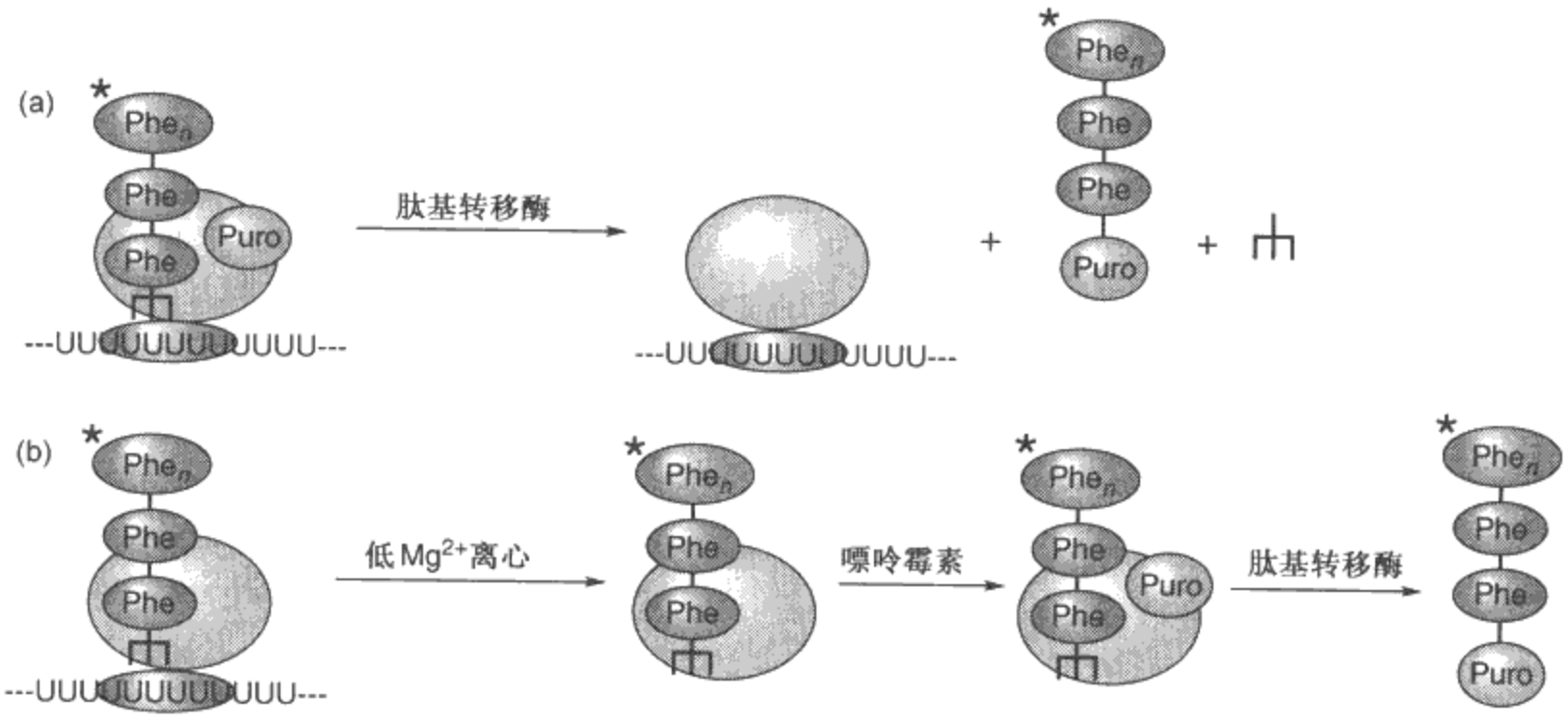


图 18.21　用于肽基转移酶分析的嘌呤霉素反应。(a) 标准嘌呤霉素反应。通过对 poly（U）信使 RNA 的翻译将标记的 poly（Phe）加载到 P 位点。然后，加入嘌呤霉素，当 poly（Phe）与嘌呤霉素间形成肽键，标记的肽酰-tRNA 被释放。(b) 只与 50S 亚基反应。同样，将标记的 poly（Phe）加载到核糖体的 P 位点，在低浓度 Mg^{2+} 缓冲液孵育，离心将 50S-poly（Phe)-tRNA 复合物与 30S 亚基和 mRNA 分离。然后，加入嘌呤霉素，通过标记肽基嘌呤霉素的释放来测定肽基转移酶活性。反应的副产品（50S 亚基和 tRNA）未在图中标出，星号表示标记的 poly（Phe)。

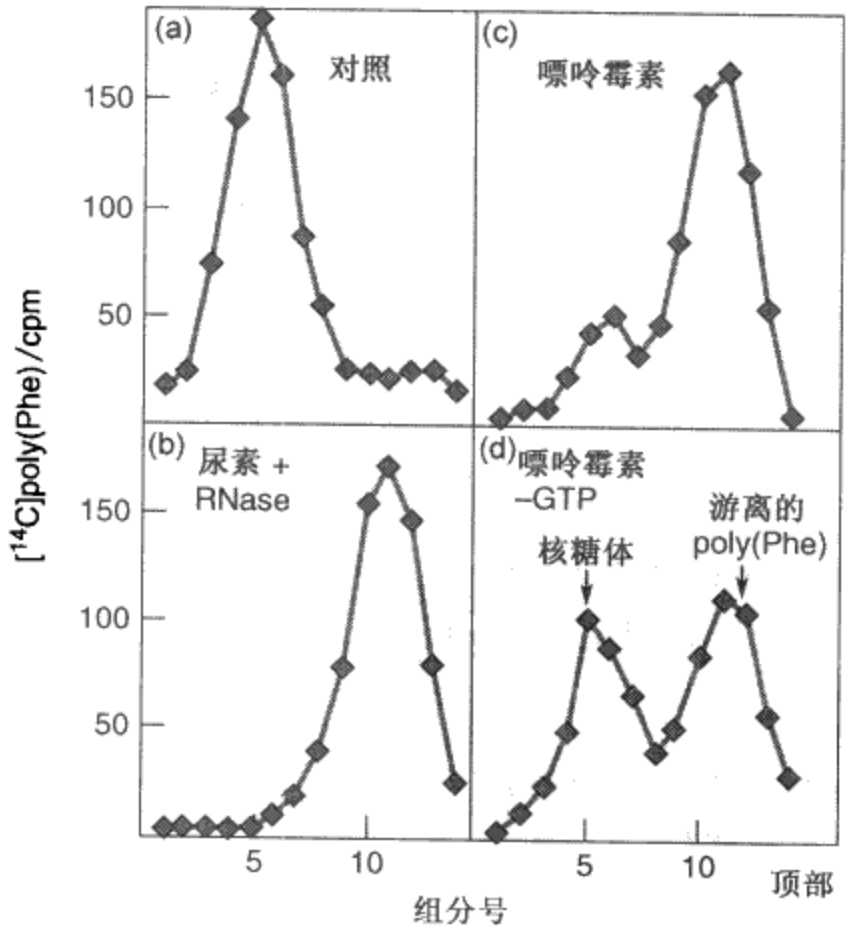


图 18.22　肽键形成的嘌呤霉素分析。Traut 和 Monro 将 [^{14}C] poly（Phe）加入核糖体中，在有或无嘌呤霉素条件下孵育，然后对产物做蔗糖密度梯度离心，分离核糖体结合的 poly（Phe）和核糖体释放的游离 poly（Phe)。对结合有 poly（Phe）的核糖体做如下处理：(a) 不处理；(b) 尿素和 RNase 处理；(c) 嘌呤霉素处理；(d) 无 GTP 时，加入嘌呤霉素。核糖体和自由 poly（Phe）的位置在 (d) 中标出。(*Source*: Adapted from Traut, R. R. and R. E. Monro, The puromycin reaction and its relation to protein synthesis. *Journal of Molecular Biology*, 10: 63-72, 1964.)

他们推测来自嗜热菌的 rRNA 结构应该非常稳定，经苯酚提取后不会变性。为了证明这一假设，他们用生活在滚烫泉水中的嗜热细菌（*Thermus aquaticus*）的 50S 亚基做了相同的实验。图 18.23 泳道 5～9 证明 *T. aquaticus* 的 50S 亚基经 3 种试剂处理后，都保留了肽基转移酶活性。

如果片段活性确实能代表肽基转移酶，那么它应该被肽基转移酶抑制剂氯霉素和碳霉素抑制。如果 rRNA 是肽基转移酶的关键因素，那么片段反应就应受 RNA 酶抑制。Noller 及同事对此两组实验条

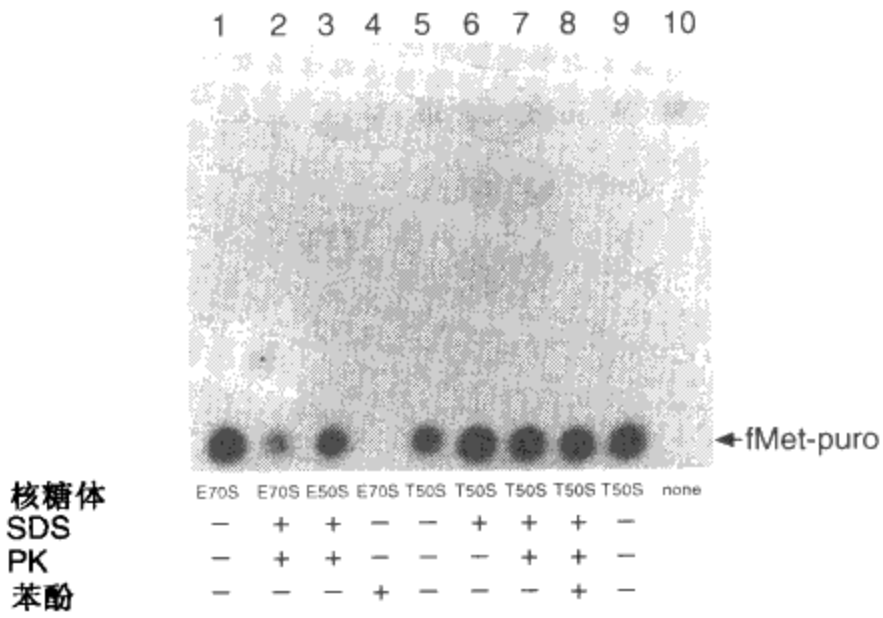


图 18.23　去蛋白质试剂对大肠杆菌和嗜热细菌核糖体肽基转移酶活性的影响。用 SDS、蛋白酶 K（PK）、苯酚单独或混合处理核糖体。然后用 CAACCA-f [^{35}S] Met 的片段反应测定所处理核糖体的肽基转移酶活性。用高压纸电泳分离 f [^{35}S] Met-嘌呤霉素，放射自显影检测。核糖体来源在底部标出：E70S 和 E50S 分别是来自*E. coli*的 70S 亚基和 50S 亚基；T50S 指嗜热细菌的 50S 核糖体亚基。标准 fMet-嘌呤霉素的位置标在图右侧标出。（*Source*：Adapted from Noller，H. F.，V. Hoffarth，and L. Zimniak，Unusual resistance of peptidyl transferase to protein extraction procedures. *Science* 256（1992）p. 1417，f. 2.）

件均获得了满意答案。对 *T. aquaticus* 完整的或处理的 50S 亚基进行片段反应，均受碳霉素、氯霉素和 RNA 酶抑制，正如所预期的那样！

那么这些实验表明 rRNA 是肽基转移酶的唯一成分了吗？Noller 及同事没有获得这个结论，部分是因为即使用最强烈的蛋白质变性剂，也不能完全去除制备物中的蛋白质。事实上，在随后与 Alexander Mankin 的合作中发现，经严格去蛋白质处理后仍有 8 个核糖体蛋白与 rRNA 结合。

Mankin、Noller 及同事将 *T. aquaticus* 的 50S 核糖体颗粒用与 Noller 原初实验相同的蛋白质破坏试剂处理，然后对剩余物质进行蔗糖梯度超速离心，发现保留肽基转移酶活性的成分以 50S 和 80S 颗粒沉降，命名为 KSP50 和 KSP80 颗粒。字母 K、S、P 分别代表蛋白酶 K、SDS、Phenol（苯酚）。接着他们检测了完整 50S 颗粒以及 KSP50 和 KSP80 颗粒，看它们都含有哪种 RNA 和蛋白质。通过凝胶电泳分离出 23S 和 5S rRNA，用双向电泳（第 5 章）来分离和鉴定颗粒中的蛋白质。难以置信的是仍然有 8 个蛋白质或多或少地保持完整，其中 4 个（L2、L3、L13、L22）以接近化学平衡的量存在，另外 4 个（L15、L17、L18、L21）的含量有所降低。通过测定各自的 N 端肽，他们再次对这 8 种蛋白质进行了鉴定，由于在两种颗粒中都出现了相同的蛋白质和 RNA，KSP80 颗粒可能是 KSP50 颗粒的二聚体。

在先前利用纯化的组分重新构成肽基转移酶的研究表明，只用 23S rRNA 和蛋白质 L2、L3、L4 就可以重构肽基转移酶活性。KSP 颗粒只缺其中的 L4。因此，考虑到重构结果和 KSP 颗粒的结果，Mankin 等认为肽基转移酶活性所需的最小成分是 23S rRNA 和蛋白质 L2、L3。

23S rRNA 在肽基转移酶中起什么作用呢？可以试图推测它具有催化作用，但是目前所得结果还不能得出这一结论。然而，Thomas Steiz 及同事在 2000 年对 50S 核糖体颗粒的 X 射线晶体学研究发现，在肽基转移酶活性中心附近没有蛋白质，只有 23S rRNA。因此，看起来 23S rRNA 确实具有肽基转移酶活性，我们将在第 19 章做详细论述。

小结　肽键是由叫做肽基转移酶的核糖体酶合成的，该酶位于 50S 核糖体颗粒上。在体外肽基转移酶活性的最小组成是 23S rRNA 和蛋白 L2、L3。X 射线晶体学研究表明 23S rRNA 位于酶的活性中心，因此看起来它在体内具有肽基转移酶活性。

延伸步骤 3：移位

一旦肽基转移酶完成它的工作，核糖体就在其 A 位点有一个肽酰-tRNA，P 位点有一个去氨酰-tRNA。下一步是移位，通过核糖体将 mRNA 和肽酰-tRNA 移动一个密码子的长度，使肽酰-tRNA 移至 P 位，并逐出氨酰-tRNA。移位过程需要延伸因子 EF-G，它能在移位完成后水解 GTP。本节内容将研究移位过程中的更多细节。

移位过程中 mRNA 的 3-核苷酸移动　首先，移位使 mRNA 准确地在核糖体上移动 3nt（一个密码子长）是合理的，任何其他长度的移动都将使核糖体进入不同的可读框，产生不正常的蛋白质产物，但证据是什么呢？Peter Lengyel 及同事在 1971 年提供了支持 3nt 假说的实验结果。他们构建了一个移位前复合物（pretranslocation complex），用的是噬菌体 mRNA、核糖体、氨酰-tRNA，但未加 EF-G 和 GTP 以便阻止移位。然后，通过加入 EF-G 和 GTP 构建了一个移位后复合物。用胰核糖核酸酶分别处理这两个复合物，消化未被核糖体保护的 mRNA，然后释放被保护的 RNA 片段并对其测序。

他们发现，片段的 3′端序列在移位前复合物是 UUU，在移位后复合物是 UUUACU。这说明移位使mRNA向左移动了 3nt，即因此 3 个核苷酸（ACU）进入核糖体受到保护。为进一步检测受保护 RNA 的 3′端序列，在释放测序之前就对其完成了翻译，发现移位前复合物中受保护 mRNA 产生的肽段以 UUU 编码的苯丙氨酸结尾，而移位后复合物中受保护 mRNA 产生的肽段以 ACU 编码的苏氨酸结尾。因此，移位使 mRNA 在核糖体上准确地移动了 3nt，即一个密码子的长度。

小结 每个移位事件使 mRNA 通过核糖体移动一个密码子的长度（3nt）。

GTP 和 EF-G 的作用 如我们在本章前面所学，*E. coli*中移位需要 GTP 和 GTP-结合蛋白 EF-G。在真核细胞中，同源蛋白 **EF-2** 执行相同的过程。1970 年，Yoshito Kaziro 及同事证明了移位对 GTP 和 EF-G 的依赖性。1974 年通过揭示移位过程何时需要 GTP，他们又有了新的发现。首先，按图 18.24 合成了一个移位底物：^{14}C 标记的 *N*-乙酰-二苯丙酰-tRNA 在 A 位点、去氨酰-tRNA 在 P 位点。该底物准备进行移位，移位可用两种方法测定：一是分析去氨酰-tRNA 从核糖体的释放，Kaziro 及同事用过滤反应混合物分析了 tRNA 的释放，并用标记的 Phe 加载来检测滤液中 tRNA 的存在。第二种分析移位的方法是检测嘌呤霉素的活性。只要移位一发生，在 P 位点标记的二肽就可与嘌呤霉素结合并被释放。表 18.5 表明，GTP 和 EF-G 单独都不能引起显著的移位，但是两者在一起就可促进移位，可通过测定去氨酰-tRNA 的释放而得知。

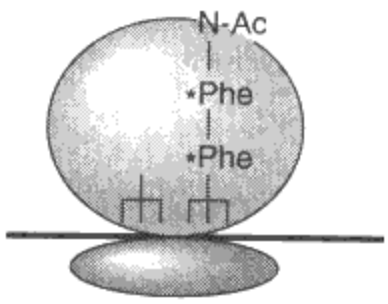


图 18.24 用于测定移位对 EF-G 和 GTP 依赖性的移位底物。Kaziro 及同事通过将 *N*-乙酰-二苯丙酰-tRNA 加载到核糖体的 A 位点、去氨酰-tRNA加载到核糖体的 P 位点，构建了移位底物。做法如下：先将核糖体和 poly（U）RNA 与 *N*-乙酰-Phe-tRNA 混合，后者占据 P 位点。然后加入通常的占据 A 位点的 Phe-tRNA。之后，肽基转移酶形成肽键，在 A 位点产生 *N*-乙酰-二苯丙酰-tRNA，在 P 位点产生去氨酰 tRNA。

在移位过程的哪一步 GTP 被水解呢？有两种主要的可能性：模型 I 要求 GTP 水解在移位前，模型 II 允许 GTP 水解在移位后。基于实验模型 II 的假设曾经较受偏爱。

表 18.5 EF-G 和 GTP 在移位中的作用

添加物	释放的 tRNA	
	pmol	Δ
实验 1		
无	0.8	
GTP	1.8	1.0
EF-G	2.4	1.6
EF-G、GTP	12.6	11.8
EF-G、GDPCP	7.5	6.7
实验 2		
无	1.6	
EF-G	1.5	0
EF-G、GTP	5.1	3.5
EF-G、GTP、梭链孢酸	6.7	5.1
EF-G、GDPCP	4.3	2.7
EF-G、GDPCP、梭链孢酸	4.7	3.1

Source: Inove-Yokosawa, N., C. Ishikawa, and Y. Kaziro, The role of guanosine triphosphate in translocation reaction catalyzed by elongation factor G. *Journal of Biological Chemistry* 249: 4322, 1974. Copyright © 1974. The American Society for Biochemistry & Molecular Biology, Bethesda, MD. Reprinted by permission.

Kaziro 及同事用 GTP 的不可水解类似物 GDPCP 做了几个实验。如果移位之前不需要 GTP，那么 GDPCP 应该像天然 GTP 那样促进移位，表 18.5 表明，GDPCP 确实产生了明显的移位，虽然没有 GTP 多。但是，研究者在使用 GDPCP 时发现必须添加化学等量（与核糖体等摩尔）的 EF-G。通常，翻译只需要催化量的 EF-G，因为 EF-G 可以反复循环。但是当 GTP 不可能水解时，如采用 GDPCP，循环就不能发生了。由此提示 GTP 水解的功能是从核糖体释放 EF-G，使 EF-G 和核糖体都能参加下一轮延伸。

实验 2 的结果在表 18.5 的下半部分，包括抗生素**梭链孢酸**（fusidic acid）的作用效果。该物质阻断 GTP 水解后 EF-G 从核糖体的释放。正常情况下，这将极大地抑制翻译，因为仅发生一次移位后就中断了翻译过程。但是在该实验中，各种处理都只能发生一轮移位，所以梭链孢酸没有发挥作用。Kaziro 及同事重复了这个实验，并用嘌呤霉素反应分析移位，获得了本质相同的结果。他们还用 GDP 代替 GTP，发现 GTP 不支持移位。

Kaziro 及同事认为 GTP 水解不是移位所绝对需要的（虽然确实有助于），因此他们推测 GTP 水解一定在移位之后。但是他们的分析需要花几分钟的

时间，远远长于移位发生所需的毫秒尺度，因而不能测定 GTP 水解和移位，区别到底谁先发生。为了严格回答这一问题，我们需要做**动力学实验**（kinetic experiment）测定前一毫秒到下一毫秒之间发生的事件。1997 年，Wolfang Wintermeyer 及同事做了这种动力学实验，明确显示 GTP 水解非常迅速并且在移位之前发生。

他们实验计划的一部分是在体外用荧光标记的肽酰-tRNA 加载到移位前核糖体的 A 位点，去氨酰-tRNA 加载到移位前核糖体的 P 位点。然后加入 EF-G-GTP 立刻测定复合物的荧光强度。利用一个**停流装置**（stopped-flow apparatus）使在毫秒（ms）尺度上进行这种动力学实验成为可能。该装置内两个或更多的溶液同时被压入一个混合室中，然后立刻进入另一个室中做分析。实验中，混合的时间仅仅是 2ms，出现初始液滴后，荧光强度迅速增加，如图 18.25a 的红色曲线。荧光强度的增加一定与移位有关，因为荧光强度被两个阻断移位的抗生素紫霉素（viomycin）和硫链丝菌肽（thiostrepton）所抑制（图 18.25a）。加入 GTP 的移位比加入不可水解 GTP 类似物或 GDP 时有效得多（图 18.25b）。

接下来，Wintermeyer 及同事比较了 GTP 水解和移位所花的时间和速度，还是用截留装置以[γ-^{32}P]GTP 测定 GTP 的水解。这一次，他们迅速将放射性 GTP 与其他成分混合，仅几毫秒之后将混合物压入另一个室内，在这里反应被高氯酸盐溶液停止。用液体闪烁计数仪测定^{32}Pi 的释放，同样还是用荧光的增强测定移位。图 18.25c 显示，GTP 水解先发生，约比移位快 5 倍。因此，Wintermeyer 及同事得出结论——GTP 水解在前并驱动移位。

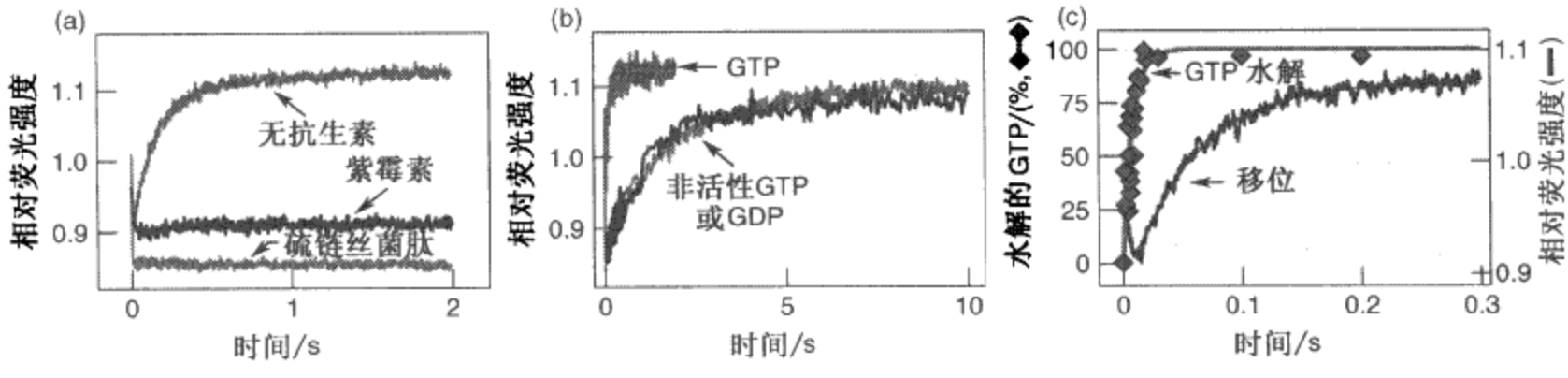


图 18.25　移位的动力学。Wintermeyer 及同事用截流动力学实验检测移位。用荧光标记的结合在 A 位上的 fMet-Phe-tRNA^{Phe}衍生物的相对荧光强度对时间（s）作图，荧光强度的增加作为移位的测定。(a) 抗生素的抗移位效果如下：红色，无抗生素；蓝色，紫霉素；绿色，硫链丝菌肽。(b) GTP 类似物的影响。将下列 GTP 类似物加入移位反应：红色，GTP；蓝色，非水解性 GTP 类似物（非活性 GTP）；绿色，GDP。(c) 测定 GTP 水解和移位的耗时。Wintermeyer 及同事利用图 (a) 和 (b) 中的截流动力学实验测定了移位反应，同时测定截流器中从 GTP 上释放的^{32}P$_i$研究 GTP 的水解情况。GTP 水解先发生，其速度比移位快 5 倍。（*Source*：Adapted from (a) Rodnina, M. V., A. Savelsbergh, V. I. Katunin, and W. Wintermeyer, Hydrolysis of GTP by elongation factor G drives tRNA movement on the ribosome. *Nature* 385 (2 Jan 1997) f. 1, p. 37. (b) f. 1, p. 37. (c) f. 2, p. 38.）

很明显，EF-G 利用 GTP 水解所释放的能量催化移位过程。这是否意味着缺乏 EF-G 就不能进行移位？实际上，某些体外条件即使缺乏 EF-G 也允许一定的移位。2003 年，Kurt Fredrick 和 Harry Noller 对此进行了目前为止最令人信服的研究，证明抗生素稀疏酶素（sparsomycin）在缺乏 EF-G 和 GTP 时，可催化移位。这一发现说明，核糖体自身即使没有 EF-G 的帮助也有能力进行移位，说明移位所需的能量储存在每个肽键形成后的核糖体、tRNA和 mRNA 组成的复合物中。

小结　尽管移位活性看起来是核糖体内在的，并可在体外无 EF-G 和 GTP 时表现出来，但是 GTP 和 EF-G 是移位所需要的。GTP 水解先于移位，并显著地促进移位。对新一轮延伸的开始，EF-G 必须依靠 GTP 水解释放的能量从核糖体上释放出来。

GTP 激酶和翻译

我们已经看到两个蛋白质 EF-Tu 和 EF-G 利用 GTP 水解来驱动翻译延伸过程重要步骤的例子。在第 17 章我们也了解到 IF2 在起始阶段起着相似的作用。在本章的最后，我们将了解到另一个因子（RF3）在翻译终止时起相同的作用。

那么所有这些过程有什么共同点呢？它们都是依靠 GTP 的能量来驱动对翻译过程非常重要的分子运动。IF2 和 EF-Tu 都能将氨酰-tRNA 带到核糖体上［IF2 转运起始氨酰-tRNA（fMet-tRNA$_f^{Met}$）到核糖体的 P 位点，而 EF-Tu 将延长的氨酰-tRNA 转运到核糖体的 A 位点］。EF-G 引发移位，其中 mRNA

和肽酰-tRNA从核糖体的A位移到P位。RF3帮助催化终止，其中连接完成的多肽与tRNA之间的肽键断开，多肽从核糖体释放。

所有这些因子都属于一大类蛋白质，叫做**G蛋白**（G protein），该类蛋白质行使范围很广的细胞功能。大多数G蛋白都具有以下特点（图18.26）。

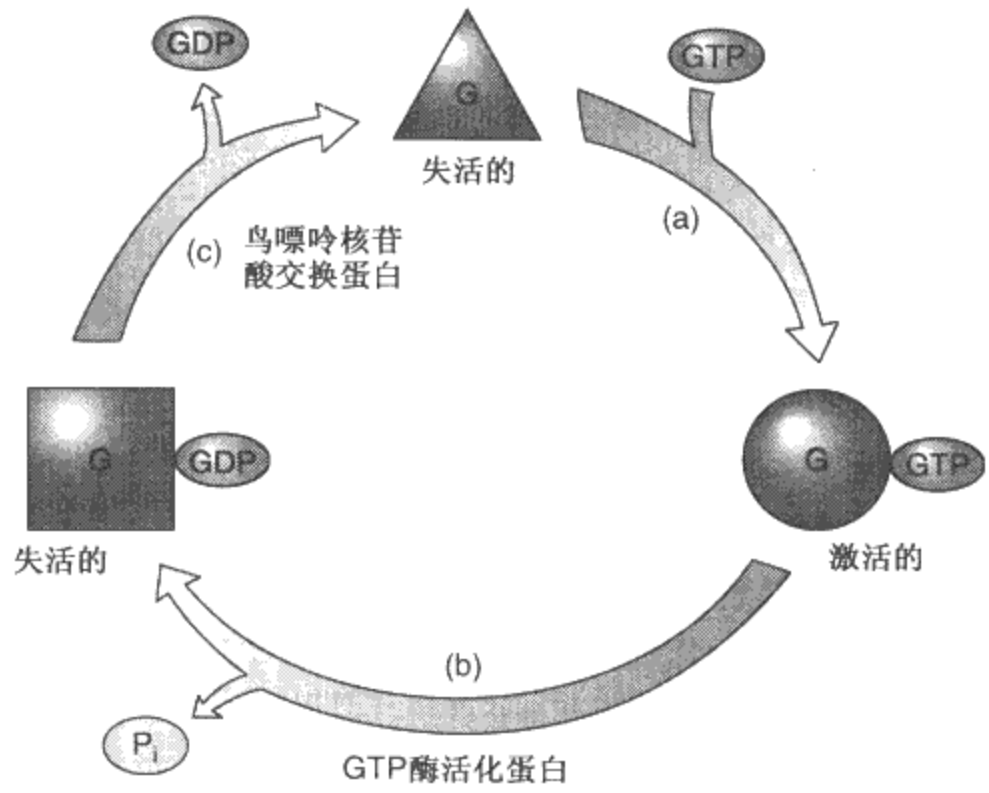


图18.26 一般化的G蛋白循环。在顶部的G蛋白（三角形）处于不与GDP或GTP结合的游离态。这种状态一般是短暂的。(a) GTP与G蛋白结合，改变其构象（变成圆形）从而激活G蛋白。(b) GTP酶活化蛋白（GAP）激活了G蛋白内在的GTP酶活性，使其将GTP水解成GDP。由此导致蛋白质构象再次发生变化，以方形表示，使G蛋白失活。(c) 鸟嘌呤核苷酸交换蛋白从G蛋白上移开GDP，使其变回原来的游离状态，等待接受另一个GTP。

1. G蛋白是GDP-结合蛋白和GTP-结合蛋白。G蛋白的“G”来自于“鸟嘌呤核苷酸”。

2. G蛋白在3种构象之间循环，取决于是否结合GDP、GTP，或不结合核苷酸，并且这些构象状态决定它们的活性。

3. G蛋白与GTP结合后才被活化行使功能。

4. G蛋白具有内在的GTP酶活性。

5. G蛋白的GTP酶活性被**GTP酶活化蛋白**（GTPase activator protein，GAP）激发。

6. 当GAP激发G蛋白的GTP酶活性时，G蛋白将自己所结合的GTP降解成GDP，并使自身失活。

7. G蛋白被**鸟嘌呤核苷酸交换蛋白**（guanine nucleotide exchange protein）重新激活，该因子从失活的G蛋白上除去GDP，允许另一分子GTP结合。鸟嘌呤核苷酸交换蛋白会立刻令人想到EF-Ts。我们已经知道EF-Ts对于用GTP替代EF-Tu上的GDP至关重要。

由于所有参与蛋白质翻译的G蛋白的GTP酶活性都受核糖体激发，因此，可以预测所有这些G蛋白的GAP应该是位于核糖体的某个（些）位点上的一个或几个蛋白质。实际上，在核糖体上已经发现了一组核糖体蛋白和部分rRNA，合起来称作**GTPase相关位点**（GTPase-associated site）或**GTPase中心**（GTPase center）。由核糖体蛋白L11及23S rRNA组成，L11是核糖体蛋白L10和L12的复合物。

小结 几个翻译因子利用GTP的能量来催化分子运动。这些因子属于庞大的由GTP激活的G蛋白家族，它们自身具有可被外部因子（GAP）激活的GTP酶活性。当它们将自身的GTP降解成GDP时就失去活性，而另外一个外部因子（鸟嘌呤核苷酸交换蛋白）用GTP替代GDP可以使之重新激活。

EF-Tu和EF-G的结构

如果EF-Tu和EF-G确实结合在核糖体的相同GTPase活性位点，那么它们应该有相似的结构，就像同一把锁的两把钥匙必须有相同的形状。对它们的X射线晶体图分析表明事实正是如此：EF-Tu-tRNA-GTP三元复合物具有与EF-G-GTP二元复合物非常相似的形状。这是合理的，因为EF-Tu与tRNA及GTP以三元复合物形式结合在核糖体上，而EF-G只与GTP以二元复合物形式结合在核糖体上。为避免GTP水解，实验者利用非水解性的GTP类似物，在EF-G中是GDP，EF-Tu-tRNA中是GDPNP。

图18.27描述了这两个复合物的三维结构。可以看到EF-G蛋白的下部（红色，右边）与EF-Tu三元复合物中tRNA（红色，左边）部分非常相似，由此使两个复合物都能结合或靠近核糖体的同一位点。

其他两个翻译因子也有核糖体依赖性GTP酶活性：原核生物的起始因子IF2（第17章）和RF3终止因子（本章后面）。由于它们也依赖于核糖体上的GTP酶活性中心，因此可以预测它们在结构上至少部分地与图18.27所示的两个复合物结构相似。

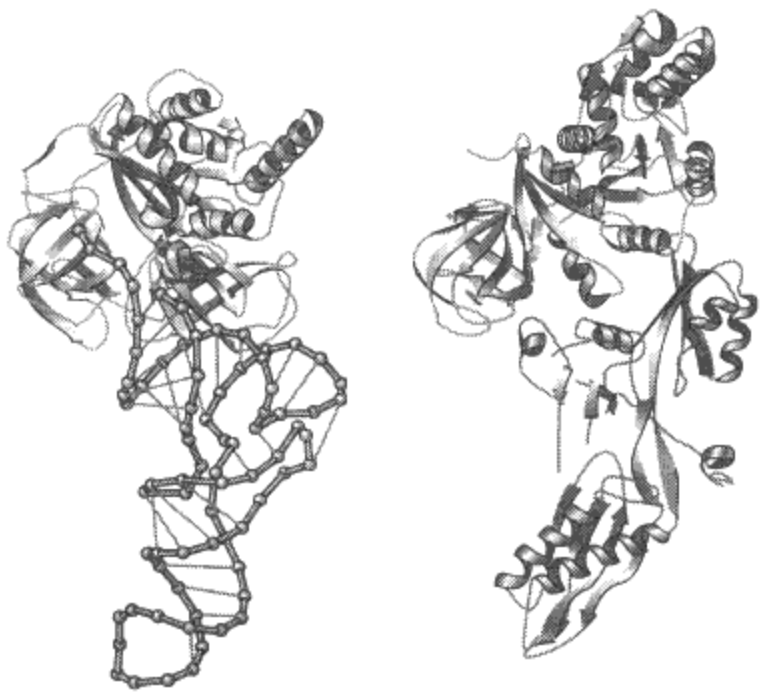


图 18.27　EF-Tu-tRNA-GDPNP 三元复合物（左）和 EF-G-GDP 二元复合物（右）三维结构的比较。三元复合物的 tRNA 部分和二元复合物的相应部位用红色标出。

另外，如果 EF-G 和 IF2 结合核糖体的相同 GTPase 中心，可以推测它们会竞争对该位点的结合。2002 年，Albert Dahlberg 及同事证明 IF2 确实与 EF-G 竞争对核糖体的结合。还有，能结合 GTPase 中心的抗生素硫链丝菌肽和微球菌素（micrococcin）也能干扰 EF-G 和 IF2 在此处的结合，因此 IF2、EF-G、EF-Tu，可能还有 RF3 都结合在至少是重叠的核糖体的 GTPase 中心。

小结　EF-Tu-tRNA-GDPNP 三元复合物和 EF-G-GDP 二元复合物的三维结构已经由 X 射线晶体图分析确定。正如预期的，它们非常相似。

18.4　翻译的终止

延伸循环不断重复，每次在多肽链上加入一个新的氨基酸。最后，核糖体遇到终止密码子，获得信号，这是翻译的最后一步——终止。

终止密码子

第一个终止密码子（**琥珀密码子**，amber codon）是 Seymour BenzerH 和 Sewell Champe 于 1962 年在 T4 噬菌体的条件突变中发现的。琥珀突变的条件性在于突变的噬菌体不能在野生型*E. coli*中复制，但可以在**抑制子**（suppressor）突变菌株中复制。*E. coli*碱性磷酸酶基因的某些突变也受这种抑制性菌株的抑制，说明碱性磷酸酶基因的突变也属于**琥珀突变**（amber mutation）。我们现在知道琥珀突变产生了终止密码子，造成翻译提前结束，产生不完全的蛋白质。这个结论的证据是什么呢？

首先，琥珀突变具有严重后果。通常的错义突变（missense mutation）最多只改变蛋白质的一个氨基酸，也可能不会对蛋白质的功能产生影响，即使蛋白质失活，通常还可用抗体检测到其存在。相反，*E. coli*碱性磷酸酶基因的琥珀突变菌株无法检测到碱性磷酸酶活性或蛋白质。这一现象符合琥珀突变引起碱性磷酸酶翻译提前终止的假说，所以没有发现全长蛋白质。

Benzer 和 Champe 进行的遗传学实验进一步验证了这一假说。他们在 T4 噬菌体的 *rIIA* 和 *rIIB* 基因间删除了一段序列，将两个基因融合在一起，见图 18.28。融合基因产生的融合蛋白具有 *B* 活性，没有 *A* 活性。接着他们在融合基因的 *rIIA* 区引入琥珀突变，该突变阻断了 *rIIB* 活性，而琥珀突变抑制子又能去除这种阻断作用。*A* 顺反子中的突变如何能抑制 *B* 顺反子的表达？它们哪个位于下游呢？很明显，琥珀突变终止翻译可以解释这一现象。如果翻译在琥珀密码子处停止，*B* 顺反子就不可能翻译。根据这个推测，琥珀突变抑制子克服了琥珀突变处的翻译终止，使翻译继续进行到 *B* 顺反子处。

Brenner 等通过对 T4 噬菌体头部蛋白基因的研究，获得了更多证明琥珀突变作为翻译终止子的直接证据。当 T4 噬菌体感染*E. coli B* 后，头部蛋白在感染后期占到总蛋白质的 50%以上，很容易被纯化。研究人员将琥珀突变引入头部蛋白基因后，很难从感染细胞中分离出完整的头部蛋白，只能分离出一些头部蛋白的片段。胰蛋白酶消化这些片段得到肽段，经鉴定为 N 端的肽段。因此，琥珀突变的头部蛋白质基因产物都是 N 端蛋白质片段。因为翻译起始于蛋白质的氨基端，这个实验证明琥珀突变引起翻译在达到羧基端之前就终止了。

琥珀突变定义了一个翻译终止密码子，而另外两个终止密码子也具有相似的彩色名字：**赭石**（ochre）和**卵白石**（opal）。赭石突变最初是通过不被**琥珀突变抑制子**（amber suppressor）抑制而鉴定出的，它们有自己的**赭石抑制子**（ochre suppressor）。类似的，卵白石突变也受**卵白石抑制子**（opal suppressor）抑制。

由于琥珀突变是由导致错义突变的突变剂产生的，我们怀疑这些突变是由普通密码子的单碱基改变而形成的终止密码子。已知遗传密码中只有 3 个不编码氨基酸的密码子：UAG、UAA 和 UGA，假定它们就是终止密码子，那么对目前所观察到的结果最简单的解释就是其中一个是**琥珀密码子**（amber codon）、一个是**赭石密码子**（ochre codon）、一个是

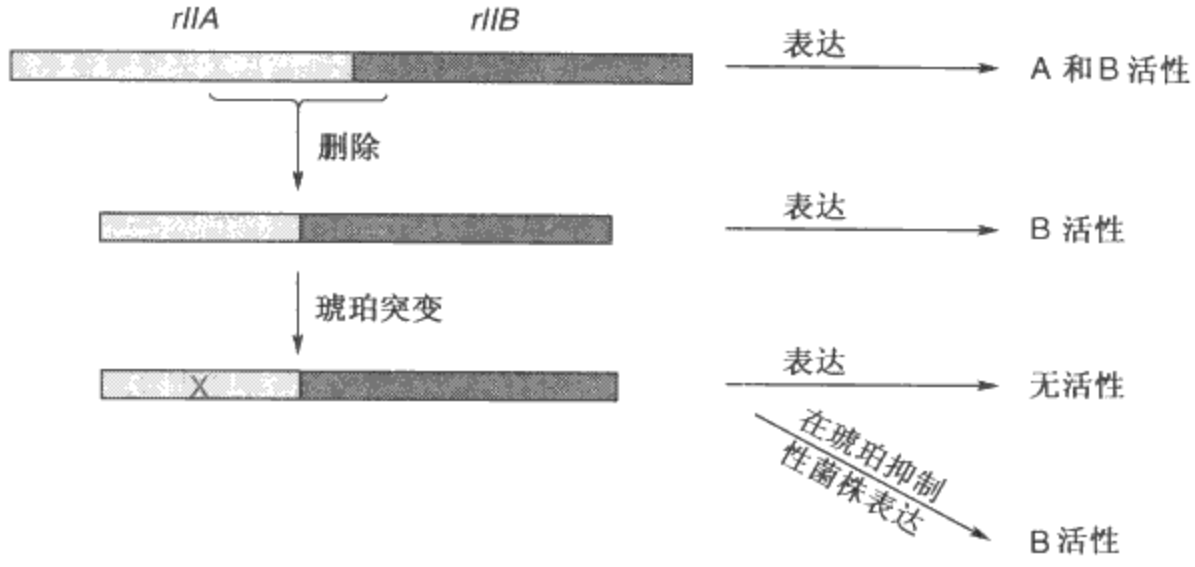


图 18.28　融合基因中琥珀突变的影响。Benzer 和 Champe 删除了括号部分的 DNA，使 *rIIA* 和 *rIIB* 顺反子融合。这个融合基因的表达产物具 *B* 活性，但没有 *A* 活性。在 *A* 顺反子中的琥珀突变使 *B* 活性失去，这种失活又可通过将融合基因转入琥珀抑制性菌株（*E. coil* CR63）而恢复。琥珀突变使 *A* 顺反子的翻译提前终止，而琥珀抑制子阻止这一终止，使融合蛋白的 *B* 部分得以表达。

卵白石密码子（opal codon），但它们的对应关系是什么呢？

1965 年，Martin Weigert 和 Alan Garen 不通过 DNA 或 RNA 测序而是蛋白质测序回答了这个问题。他们研究了 *E. coli* 碱性磷酸酶基因一个位点的琥珀突变，在野生型细胞中这个位点是色氨酸，其唯一密码子是 UGG。因为琥珀突变始于单碱基改变，我们也知道琥珀密码子是通过单碱基改变与 UGG 相关的，要找出是哪个碱基发生了改变，Weigert 和 Garen 将几个不同回复突变子插入这个位点，然后测定氨基酸序列。回复突变子应该是琥珀密码子的单碱基改变产生的，其中一些在关键位点上是色氨酸，但大多数是其他氨基酸，如丝氨酸、酪氨酸、亮氨酸、谷氨酸、谷氨酰胺酸和赖氨酸。这些氨基酸可以很好地替代色氨酸，产生碱性磷酸酶活性。我们的难题是要推出一个密码子，其单碱基改变至少要对应这些氨基酸中的一种，包括色氨酸。图 18.29 证明 UAG 是问题的答案，因此，这个密码子一定是琥珀密码子。

按照相同的逻辑，包括琥珀突变体可通过单碱基改变而突变为赭石突变体这一现象，Sydney Brenner 及同事推断赭石密码子一定是 UAA。Severo Ochoa 及同事证实 UAA 是一个终止信号，他们用合成的信使 $AUGUUUUAAA_n$ 来指导二肽 fMet-Phe 的合成和释放（AUG 编码 fMet，UUU 编码 Phe，UAA 则是终止信号）。由于 UAG 和 UAA 分别是琥珀密码子和赭石密码子，通过排除法可知 UGA 一定是卵白石密码子。现在，我们有了数千个基因的碱基序列，足以清楚地看出这 3 个密码子确实是作为终止信号的。有时甚至可以发现两个终止密码子并列在一起（如 UAAUAG），这提供了一个故障保护终止信号，即使一个终止密码子被抑制，终止也能实现。

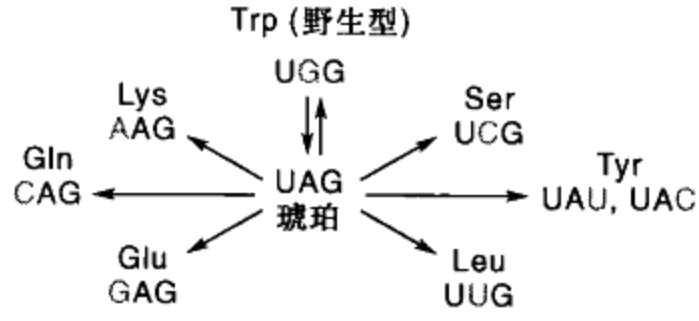


图 18.29　琥珀密码子是 UAG。琥珀密码子由色氨酸密码子（UGG）单碱基突变而来。下列氨基酸替代色氨酸使基因回复到功能状态：丝氨酸、酪氨酸、亮氨酸、谷氨酸、谷氨酰胺或赖氨酸。紫红色表示在所有这些回复突变子中被改变的单个碱基，包括编码色氨酸的野生型回复突变子。

小结　mRNA 分子中的琥珀、赭石和卵白石突变分别产生 UAG、UAA 和 UGA 3 个终止密码子，因此造成肽链翻译的提前终止。这 3 个密码子也是 mRNA 编码区末端的正常终止信号。

终止密码子的抑制

抑制子如何克服未成熟终止信号的致死效应呢？1965 年，Mario Capecchi 和 Gary Gussin 证明 *E. coli* 抑制子菌株的 tRNA 能抑制 R17 噬菌体外壳顺反子的琥珀突变，由此将 tRNA 鉴定为抑制分子，但它是如何起作用的呢？Brenner 及同事在对抑制子 tRNA 测序时发现了答案。他们将琥珀抑制子 tRNA 的基因导入 ϕ80 噬菌体，并用该重组噬菌体感染在 *lacZ* 基因上有琥珀突变的 *E. coli*。由于抑制子 tRNA

的存在，被感染细胞通过插入一个酪氨酸来替代终止，抑制了琥珀突变。当 Brenner 及同事测序这个抑制子 tRNA 时，发现只有一个碱基不同于野生型的 $tRNA^{Tyr}$：反密码子的第一个碱基由 G 变为 C，如图 18.30 所示。

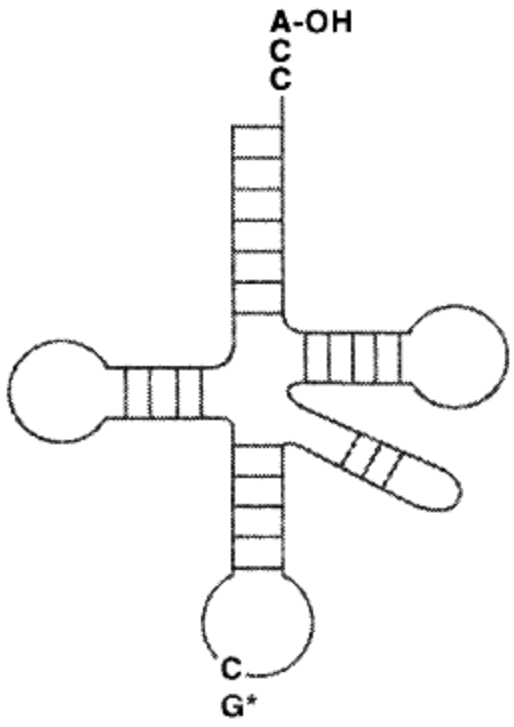


图 18.30　野生型 *E. coli* 的 $tRNA^{Tyr}$ 和 *E. coli* 琥珀突变抑制子 tRNA 的序列比较。野生型 $tRNA^{Tyr}$ 中的 G*，被抑制子 tRNA 中的 C 所取代。（*Source*：Adapted from Goodman, H. M., J. Abelson, A. Landy, S. Brenner, and J. D. Smith, Amber suppression: A nucleotide change in the anticodon of a tyrosine transfer RNA. *Nature* 217: 1021, 1968.）

图 18.31 说明了这种改变的 tRNA 是如何抑制琥珀密码子的。我们从编码谷氨酰胺（Gln）的密码子 CAG 开始，它与 $tRNA^{Gln}$ 上的反义密码子 3′-GUC-5′配对。假设 CAG 密码子突变成 UAG，那么就不再与 $tRNA^{Gln}$ 配对，而是吸引终止机器停止翻译。现在，第二个突变在 $tRNA^{Tyr}$ 上的反密码子发生，由 AUG 变为 AUC（同样按 3′→5′阅读）。这个新产生的 tRNA 是个抑制子 tRNA，其反密码子与琥珀密码子 UAG 互补，能与 UAG 终止密码子配对并且把酪氨酸插入到延伸的肽链上，使核糖体通过终止密码子而不终止翻译。

小结　大部分抑制子 tRNA 具有改变的反密码子，能够识别终止密码子，通过插入一个氨基酸允许核糖体通过终止密码子移动到下一个密码子而阻止终止。

释放因子

由于终止密码子和普通的密码子一样都是三联体，我们可能期望这些终止密码子会像其他密码子那样被 tRNA 解码。但是 1967 年 Mario Capecchi 证明，终止密码子不能常规性地被 tRNA 识别，而是由称为**释放因子**（release factor，RF）的蛋白质所识别。他设计了以下实验来鉴定释放因子（图 18.32），用*E. coli*核糖体加 R17 噬菌体 mRNA，将其外壳顺反子的第 7 个密码子突变为 UAG（琥珀），在这个琥珀密码子之前是编码苏氨酸的 ACC。将核糖体与 mRNA 在无苏氨酸条件下共孵育，以便产生一个五肽，并在苏氨酸密码子处停止。然后提取带有五肽的核糖体，把它们放入只含 EF-Tu、EF-G（与核糖体结合的）和 [^{14}C] Thr-tRNA 的系统中，核糖体将带有标记的苏氨酸整合到肽段上，在 P 位点产生一个带标记的六肽，准备好释放。为找到释放因子，Capecchi 不断添加核糖体的上清液组分，直到标记的多肽被释放。最终，他发现了这个因子，并将其称为释放因子（RF），RF 是一种蛋白质而不是 tRNA。

Nirenberg 及同事设计了一种更简单的技术（图 18.33），该方法是在本章前面介绍过的他们鉴定密码子方法的翻版。他们用核糖体、三联体 AUG 和 [^{3}H] fMet-$tRNA_f^{Met}$ 构成三元复合物。起始密码子和氨酰-tRNA 结合在复合物的 P 位点，所以标记的氨基酸等待释放。将复合物与释放因子的粗提物及任一终止密码子（UAG、UAA 或 UGA）一起培养，使标记的 fMet 释放。在分析中，终止的三核苷酸进入 A 位点，并且如有合适的释放因子，就下令释放。表 18.6 显示，释放因子 1（**RF1**）与终止密码子 UAA 和 UAG 合作，引起 fMet 的释放，而释放因子 2（**RF2**）与 UAA 和 UGA 合作。接下来的研究显示，UAA 或 UAG 能够指导纯化的 RF1 对核糖体的结合，而 UAA 或 UGA 指导 RF2 的结合。这进一步说明，这些 RF 能够识别特定的翻译终止信号。第三个释放因子（**RF3**）是一个核糖体依赖性 GTPase，结合 GTP 并帮助其他两个释放因子结合到核糖体上。基于 EF-G 对结合了 tRNA 的 EF-Tu 结构的模仿，Jens Nyborg 及同事推测 RF3 的结构应该与 EF-Tu-tRNA-GTP 三元复合物的蛋白质部分类似，而 RF1 和 RF2 应模仿 tRNA 的结构。RF1 和 RF2 与 tRNA 竞争核糖体上的结合位点，像 tRNA那样识别密码子，并且与 tRNA 大小相似，这些事实都与该假设一致。

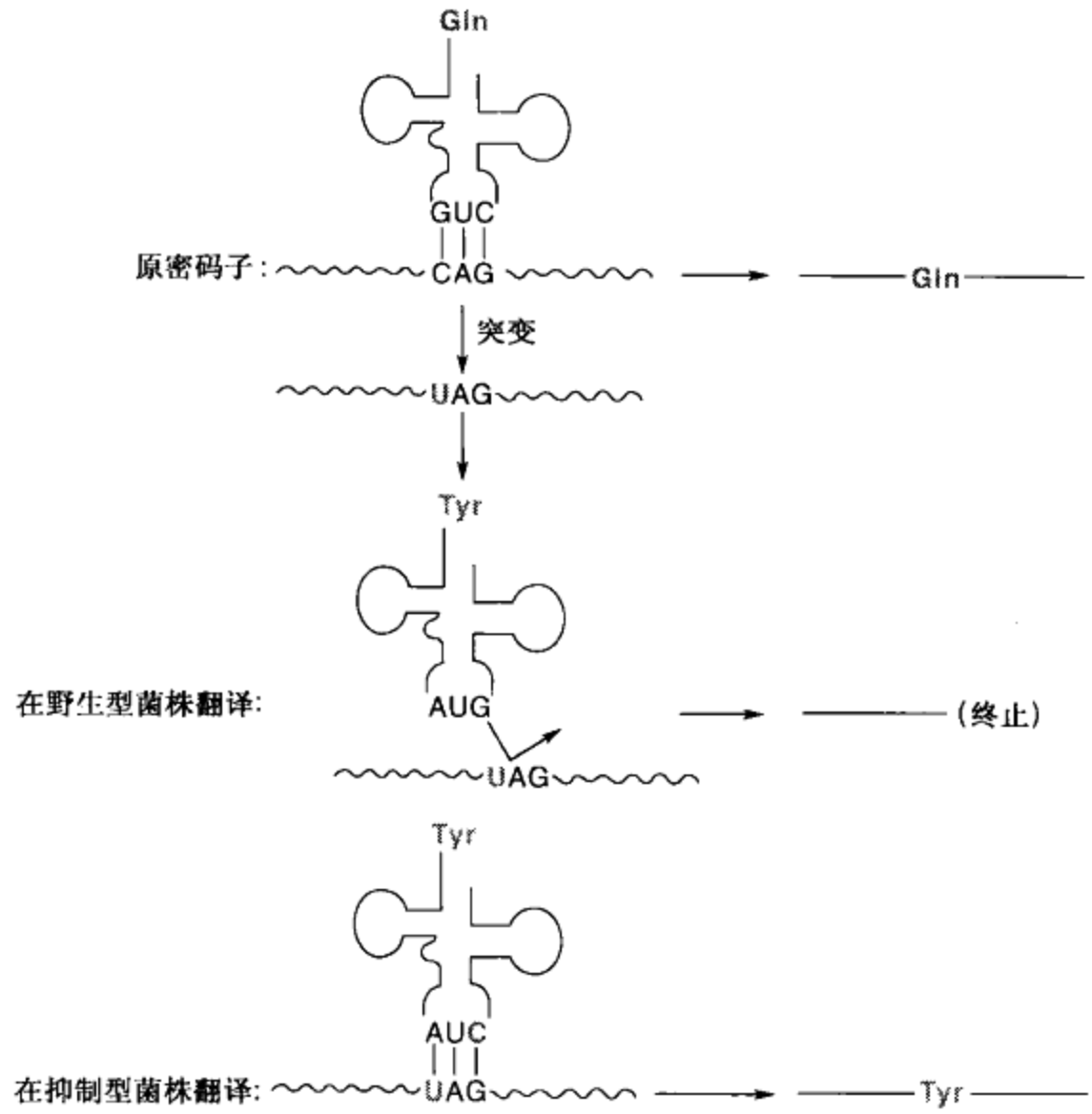


图 18.31　抑制机制。上图：野生型*E. coli*基因中最初的密码子是CAG，可以被 $tRNA^{Gln}$ 识别。中图：这个密码子突变为 UAG，可被野生型*E. coli*翻译为终止密码子。$tRNA^{Tyr}$ 的反密码子（AUG），不能翻译琥珀密码子。底图：一个抑制性菌株含有突变的酪氨酸 tRNA，反密码子 AUG 突变为 AUC。这个改变的反密码子能识别琥珀密码子并引起酪氨酸（灰色）插入而不是终止。

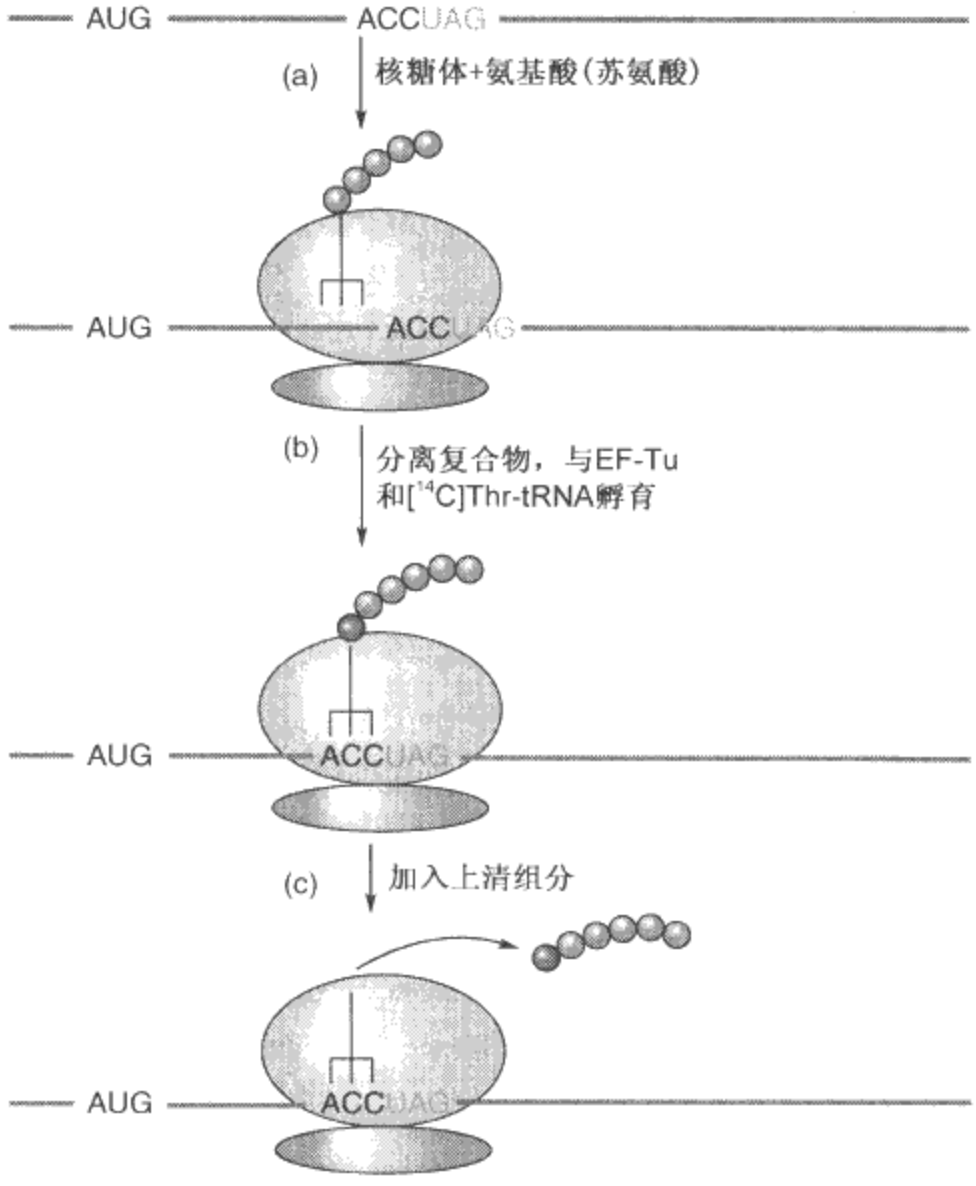


图 18.32　释放因子检测。Capecchi 以突变的 R17 噬菌体 mRNA 开始其 UAG 终止密码子（蓝色）跟在第 6 个编码苏氨酸的密码子 ACC 之后。（a）加入除苏氨酸之外含所有氨基酸的翻译系统，这样允许一个五肽的合成（绿色）；（b）分离复合物并和 EF-Tu、[^{14}C] Thr-tRNA 共孵育，形成一个包含 [^{14}C] 苏氨酸（红色）的六肽，准备释放。（c）最后，加入含释放因子的上清组分，释放标记的六肽。

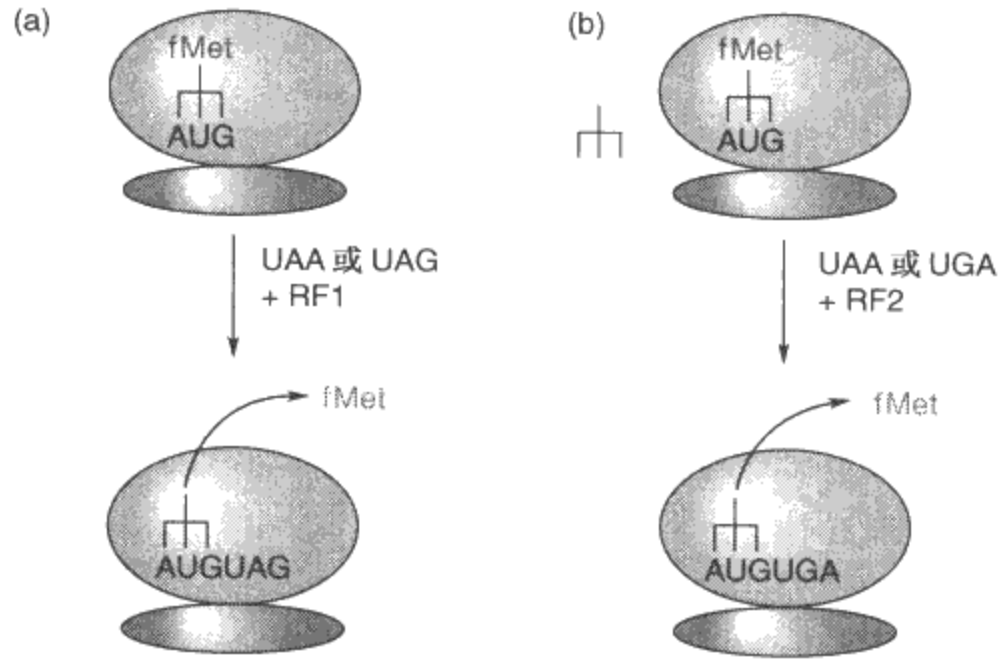


图 18.33　Nirenberg 对释放因子的分析。Nirenberg 在核糖体的 P 位点加载起始密码子 AUG 和 [^{3}H] fMet-tRNA$_f^{Met}$。然后加入一个终止密码子和一个释放因子，释放出标记的 fMet。(a) RF-1 与 UAA 或 UAG 作用；(b) RF-2 与 UAA 或 UGA 作用。

表 18.6　RF1 和 RF2 对终止密码子的响应

添加物		下列条件时 pmol[^{3}H]fMet 的释放	
释放因子	终止密码子	0.012mol/L Mg^{2+}	0.030mol/L Mg^{2+}
RF1	无	0.12	0.15
RF1	UAA	0.47	0.86
RF1	UAG	0.53	1.20
RF1	UGA	0.08	0.10
RF2	无	0.02	0.14
RF2	UAA	0.22	0.77
RF2	UAG	0.02	0.14
RF2	UGA	0.33	1.08

Source: From "Release Factors Differing in Specificity for Terminator codons," by W. Scolnick, R. Tompkins, T. Caskey, and M. Nirenberg, Proceedings of the National Academy of Sciences, USA, 61: 772, 1968. Reprinted with permission of the authors.

那么真核生物的释放因子又是怎样的呢？第一个真核释放因子(eRF) 于 1971 年通过与 Nirenberg 相似的方法被发现。1994 年，Lev Kisselev 小组仍然利用 Nirenberg 的方法最终纯化了 eRF，且成功地克隆和测序了 eRF 基因。克隆和测序基因采用广泛使用的方法。采用与 Nirenberg 相似的方法通过检测 fMet 的释放来测定 eRF，该小组纯化出的有活性 eRF 在 SDS-PAGE 电泳图谱上只出现一条主带，然后用双向电泳将其与其他蛋白质分离。切下含 eRF 的胶，并用胰蛋白酶消化，对其中 4 个水解肽段进行**微测序**（microsequencing)。所得序列与人、非洲爪蟾、酵母、拟南芥蛋白质序列高度相似。因此，可以用已克隆的 *Xenopus* 基因（C11）作为探针从人 cDNA 文库中寻找相应基因。为证实克隆的 *Xenopus* 基因（C11）和人基因（TB3-3）具有 eRF 活性，他们在细菌或酵母中表达了这些基因，并用四核苷酸（其中有些包含终止密码子）的 fMet 释放实验进行检测。表 18.7 显示两种蛋白质在有终止密码子情况下，都能从核糖体上释放 fMet。*Xenopus* 蛋白表达时带有寡聚组氨酸标签，因此以带有组氨酸标签的不相关蛋白质作为阴性对照。结果发现，用抗 C11 的抗体可以阻断其释放因子的活性，但不相关抗体 (anti-Eg5) 不能。

表 18.7 显示，eRF 能识别全部 3 个终止密码子，而不像原核细胞的两种释放因子都只能识别 2 个。那么 eRF 也像原核细胞 RF1 和 RF2 一样，与 G

表 18.7　纯化的非洲爪蟾 C11 和人 TB3-1 的多肽释放活性

蛋白质	使用的四核苷酸	释放的 f[^{35}S] Met/pmol
带组氨酸标签的 C11	UAAA	0.38
	UGAA	0.39
	UAGA	0.35
	poly(A、G、U)	0.28
	UGGA	0
带组氨酸标签的人 prothymosine	UAAA	0
带组氨酸标签的 HIV	UAAA	0
反转录酶 TB3-1	UAAA	0.20
	UGAA	0.23
	UAGA	0.17
	UGGA	0

Source: Adapted from Frolova, L., X. Le Goff, H. H. Rasmussen, S. Cheperegin, G. Drogeon, M. Kress, I. Arman, A. L. Haenni, J. E. Celis, M. Philippe, J. Justesen, and L. Kisseler, A highly conserved eukaryotic protein family possessing properties of polypeptide chain release factor, *Nature* 372:701, 1994.

蛋白协作吗？1995 年，Michel Philippe 及同事得出了肯定的答案。他们在 *X. laevis* 细胞中发现了一种与原核释放因子 RF3 相对应的真核释放因子 **eRF3**。eRF3 家族的另一成员，酵母 Sup35 蛋白有一个鸟嘌呤核苷酸结合结构域，其对酵母生长是必需的。因此，真核细胞中包含两种释放因子——eRF（重命名为 **eRF1**）和 eRF3。eRF1 具有与 RF1 和 RF2 相同的作用，eRF3 与细菌 RF3 的作用相同。

David Barford 及同事获得了 eRF1 的晶体结构，进一步证实释放因子是 tRNA 的分子模仿——这本是对原核释放因子提出的设想。如图 18.34 所示，eRF1 的结构与 tRNA 的双末端结构极为相似。其中一端有一个类反义密码子区，可识别全部 3 个终止密码子。在另一端，应是氨基酸或多肽与 tRNA 结合的部位，存在有释放因子的活性位点，该区域具有稳定的 Gly-Gly-Gln 基序，可能参与成熟肽链和 tRNA 间肽键的水解。与这个假说一致的是代换 Gly-Gly-Gln 基序上的任一氨基酸，会使蛋白质完全失活。

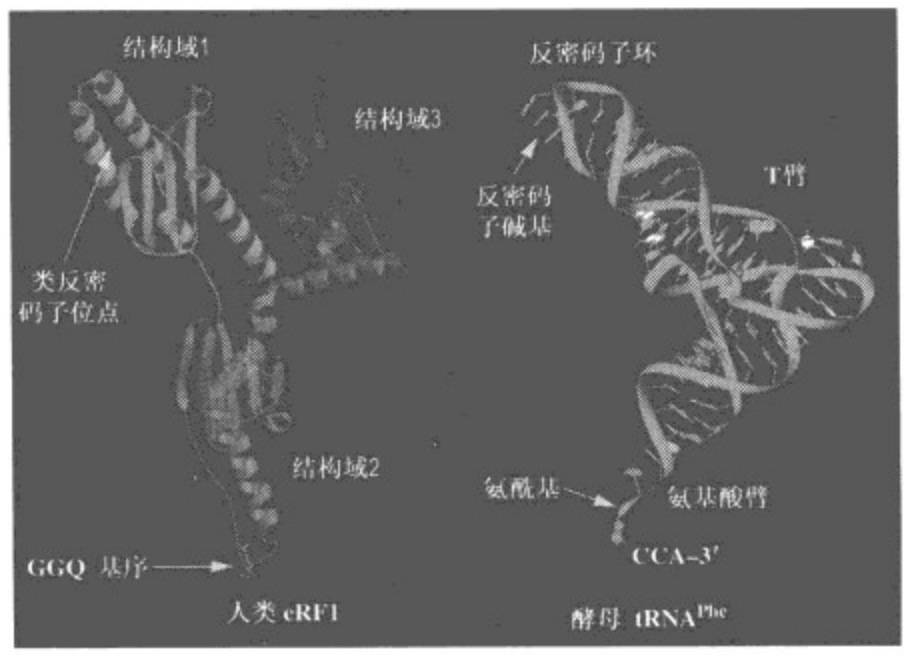


图 18.34　eRF1 结构模拟图。人 eRF1 和酵母 $tRNA^{Phe}$ 晶体结构比较，eRF1 的 1、2、3 结构域的位置分别与 tRNA 的反义密码子环、氨基酸臂（受体臂）和 T 臂相对应。eRF1 的 GGQ 结构与氨酰-tRNA 上酰基化基团处于相似的位置，因此能够参与邻近的核糖体 P 位肽酰-tRNA上的 tRNA 和多肽之间的酯键水解。（*Source*：Song，H. P.，Mugnier，A. K. Das，H. M. Webb，D. R. Evans，M. F. Tuite，B. A. Hemmings，and D. Barford，The crystal structure of human eukaryotic release factor eRF1-mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. *Cell* 100（4 Feb 2000）f. 6，p. 318. Reprinted by permission of Elsevier Science.）

值得注意的是，Nakamura 及同事发现，eRF1 上的类反义密码子基序（anticodon-like motif）与早先 Yoshikazu Nakamura 及同事鉴定的 RF1 和 RF2 上的相应基序完全不同，这种差异性与 eRF1 的序列不同于 RF1 或 RF2（除了 Gly-Gly-Gln 基序）是一致的，尽管原核释放因子之间都非常类似。因此，真核细胞（及古菌）和原核细胞的释放因子可能具有不同的起源，并且由此进化出识别终止密码子的不同方式。

此外，虽然 RF2 的晶体结构与 tRNA 极其相似（甚至比 eRF 与 tRNA 的相似程度还高），负责转换终止密码子特异性（UAG 对 UGA）的三核苷酸却远离该分子在结构上模仿 tRNA 反密码子的部位。因此，如果 RF2 分子如其结构所提示的那样进入核糖体的 A 位点，则三核苷酸就不能通过直接识别终止子来替换反密码子。还有，现在甚至还不清楚 RF2 是否能够如其类 tRNA 结构所提示的那样进入核糖体。不得不说这个漂亮的仿真模型还存在一些严重瑕疵是令人失望的，我们也还不知道终止密码子的识别是怎样发生的。

小结　原核生物翻译终止受 3 种因子的介导，即 RF1、RF2 和 RF3。RF1 识别终止密码子 UAA 和 UAG；RF2 识别 UAA 和 UGA；RF3 是一个 GTP 结合蛋白，能促进 RF1 和 RF2 与核糖体结合。真核细胞有两个释放因子：eRF1 识别所有 3 个终止密码子；eRF3 是一种核糖体依赖性 GTP 酶，能帮助 eRF1 释放翻译成熟的肽链。

异常终止的处理

两类异常 mRNA 可导致翻译的异常终止。一类是我们已经看到的能引发提前终止的“无义”突变。第二类是有些 mRNA（**非终止 mRNA**，non-stop mRNA）缺乏终止密码子，可能是 mRNA 合成时在终止密码子的上游就停止了，核糖体翻译时通过了这些非终止 mRNA 然后停滞。这两种情况都给细胞带来了问题，无论是提前终止翻译还是停滞的核糖体都会产生不完整的蛋白质，对细胞产生不利影响。停滞核糖体带来的另一个问题是它失去作用，不再参与进一步的蛋白质合成。我们先看一下细胞处理非终止 mRNA 的几种方式，然后了解提前终止产物的降解机制。

非终止 mRNA。要处理非终止 mRNA，细胞需要降解异常蛋白质产物，释放核糖体亚基，以便它们能够参与翻译而不是永远保持停滞状态。原核生

物和真核生物这一过程的机制不尽相同。原核生物利用**转移信使 RNA**（transfer-messenger RNA，**tmRNA**）来恢复停滞的核糖体，并标记需降解的非终止 mRNA（**tmRNA 介导的核糖体恢复**）（tmRNA-mediated ribosome rescue）。tmRNA 长约 300nt，其 5′端和 3′端接合在一起形成一个**类 tRNA 结构域**（tRNA-like domain，**TLD**），与 tRNA 相似（图 18.35）。实际上这种相似性非常强，以至于 tmRNA 能够加载丙氨酸。一旦加载，丙氨酰-tmRNA 就能与核糖体的 A 位点结合，经肽基转移酶催化，可以将其丙氨酸安装到停滞的多肽链上。

肽基转移酶反应之后，tmRNA 的中心部位开始发挥作用（图 18.36）。这部分 tmRNA 含有一个短的可读框（ORF），它处于 A 位点，这样核糖体就从翻译非终止 mRNA 转换成翻译 tmRNA，此过程称为**反式翻译**（*trans*-translation）。tmRNA 的可读框编码一个短的疏水肽，疏水肽将被加到停滞多肽的羧基末端。该短肽瞄准整个有待降解的多肽，减小其破坏细胞的能力。

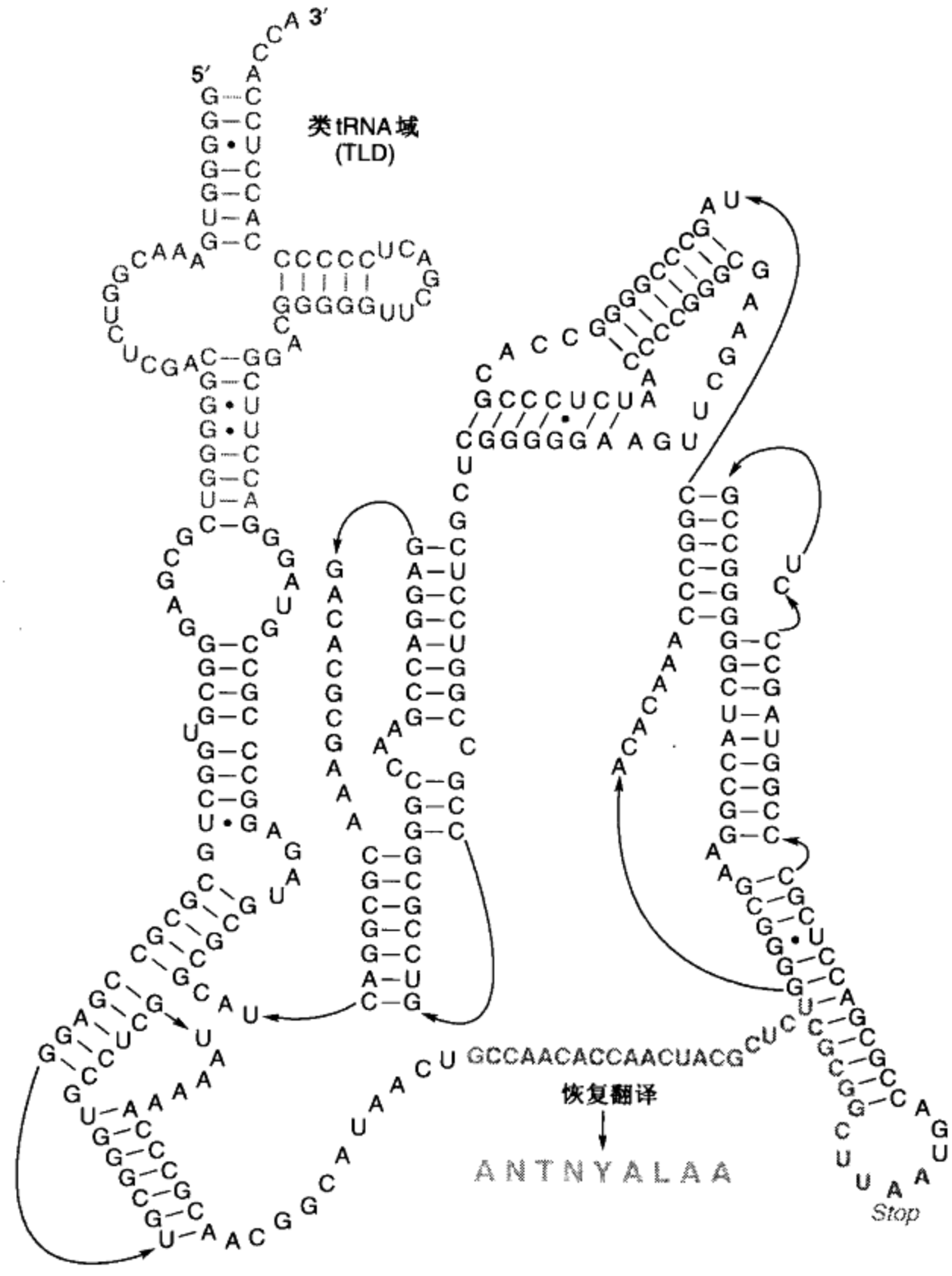


图 18.35　嗜热细菌 tmRNA 的结构。TLD 在左上方红色部分，ORF 在底部蓝色标出。ORF 编码的肽用黄色标出。（*Source*：Adapted from Valle et al.，Visualizing tmRNA entry into a stalled ribosome，*Science* 300：128，fig. 1，2003.）

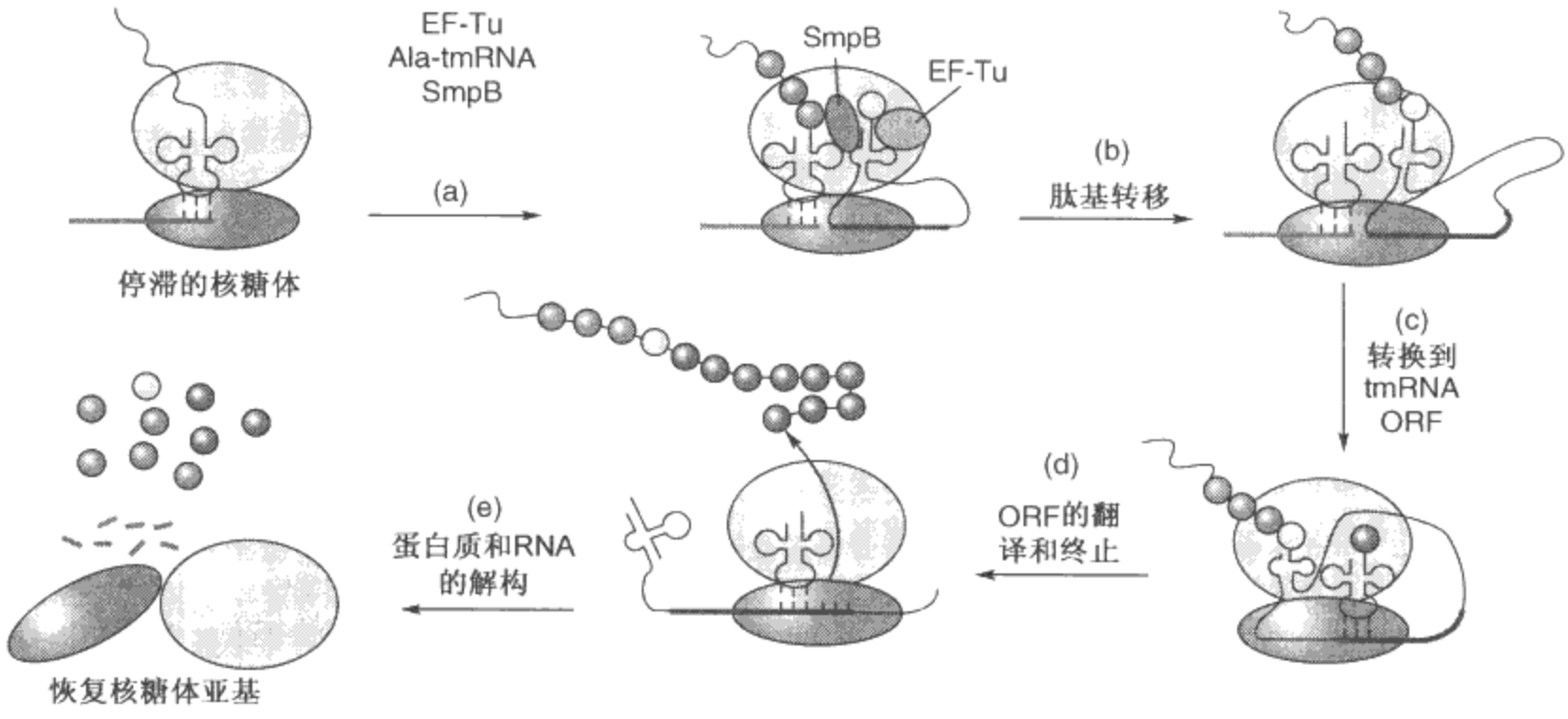


图 18.36　tmRNA 介导的非终止 mRNA 和多肽的释放机制。(a) EF-Tu、丙氨酰 tmRNA 和 SmpB 结合到停滞在非终止 mRNA（棕色）的核糖体 A 位点，SmpB 帮助有 tRNA 类似结构域的 tmRNA 结合到核糖体上；(b) 核糖体的肽基转移酶将丙氨酸（黄色）从 tmRNA 转到停滞的多肽上（绿色）；(c) 核糖体转而阅读 tmRNA 的 ORF（紫色）；(d) 核糖体完成 tmRNA 的 ORF 翻译，在停滞多肽上增加了 9 个氨基酸（红色），并释放多肽；(e) 合起来，这些额外的氨基酸标记整个多肽以便水解。同时，非终止mRNA 被降解，可能是由与 tmRNA 结合的 RNase R 所水解。(*Source*: Adapted from Moore, S. D., K. E. McGinness, and R. T. Sauer, A glimpse into tmRNA-mediated ribosome rescue. *Science* 300 [2003] p. 73, f. 1.)

很明显，tmRNA 并不完全像 tRNA。一是它缺乏反密码子，因而没有密码子-反密码子配对。我们知道在校正时，密码子-反密码子配对对于避免氨酰-tRNA 解离非常重要。tmRNA 和真正 tRNA 的第二点不同是 tmRNA 没有标准的 D 环，而是通过 **SmpB** 蛋白解决这些问题。2003 年，Joachim Frank 和 V. Ramakrishnan 获得了嗜热菌（*T. thermophi-lus*）的 EF-Tu、tmRNA 和 SmpB 复合体结合到核糖体上的低温电子显微图像。这一研究表明 SmpB 与 tmRNA 和 EF-TU 结合，建立了与核糖体的联系，而正常情况下是通过 RNA 的 D 环联系的。因此，尽管 tmRNA 自身缺乏能够与核糖体紧密结合的成分，但 SmpB 能够将其结合到核糖体上。

核糖体被 tmRNA 释放后，对非终止 mRNA 会有什么影响呢？这个答案还不确定。但 tmRNA 能够与一个 3′→5′外切核酸酶 **RNase R** 同时被提纯，由此产生一个很有吸引力的假说：在非终止 mRNA 与新核糖体形成复合物之前，RNase R 就将其降解了。

真核细胞没有 tmRNA，那么它们是怎样处理非终止 mRNA 的呢？图 18.37 阐述了目前的假说，停滞在非终止 mRNA 末端的核糖体 A 位点会含有 poly (A) 尾部的 0～3 个核苷酸，这种状态被 **Ski7p** 蛋白的羧基末端结构域所识别，该结构域类似于延伸因子 EF1A 和终止因子 eRF3 的 GTPase 结构域，这些结构域通常与核糖体 A 位点结合，Ski7p 也是这样的。另外，Ski7p 还与细胞质的**外来体**（exosome）（细胞核外来体的同源物）紧密结合。这两种外来体都能降解 RNA，Ski7p-外来体复合物然后重组 **Ski 复合物**（Ski complex）到核糖体 A 位点，并与非终止 mRNA 的末端临近。最后，外来体降解非终止 mRNA。由非终止 mRNA 产生的异常蛋白质的命运还不清楚。

小结　原核生物通过 tmRNA 介导的核糖体恢复来处理非终止 mRNA。类似于丙氨酰-tRNA 的丙氨酰-tmRNA 结合到停滞在非终止 mRNA 上的核糖体的空 A 位上，将其丙氨酸添加到停滞的多肽上。然后，核糖体转而翻译 tmRNA 的 ORF，在终止前给多肽再添加 9 个氨基酸。这些额外的氨基酸标记了有待降解的多肽，同时核酸酶降解非终止 mRNA。真核生物的核糖体在非终止 mRNA 的 poly (A) 尾末端将 Ski7p-外来体复合物吸引到空的 A 位上。接着，Ski 复合物结合到 A 位，外来体正好定位于非终止 mRNA 的末端，降解 RNA，估计异常多肽也被降解了。

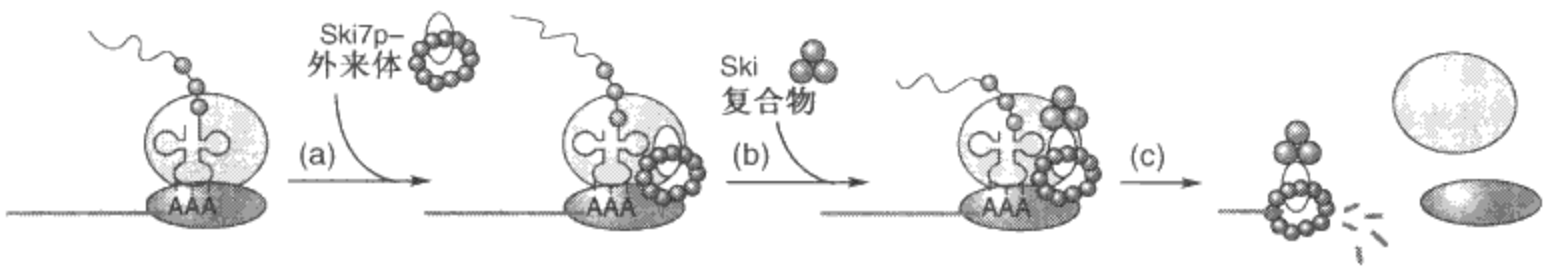


图 18.37　外来体介导的真核生物非终止 mRNA 的降解模型。（a）停滞在非终止 mRNA 末端的核糖体 A 位点含有 0～3 个 poly（A）尾的核苷酸。这里，A 位点上没有成串的 A。核糖体的此种状态吸引了 Ski7p-外来体复合物与空 A 位结合。（b）接着，Ski 复合物结合在 A 位。（c）引发了非终止 mRNA的降解和核糖体亚基的释放。

提前终止　带有提前终止密码子（无义密码子）的信使 RNA 也会产生对细胞具有潜在危害的异常不完整蛋白质产物。真核细胞演化出了两条途径解决这一问题（图 18.38）：**无义介导的 mRNA 降解**（nonsense-mediated mRNA decay，**NMD**）和**无义相关的修饰剪接**（nonsense-associated altered splicing，**NAS**）。

NMD 依赖于对提前终止密码子的识别。很显然，在每个 mRNA 的末端有一个真正的终止密码子，细胞必须以某种方式区别真正的终止密码子和提前终止密码子。一种假说是，真核细胞通过测量终止密码子和**下游不稳定元件**（downstream destabilizing element）之间的距离进行区别，如果距离较短，该终止密码子可能是真的，如果距离较长，则该终止密码子可能是提前的。

哺乳动物的下游不稳定元件是一组蛋白质的集合体，叫做**外显子交界复合物**（exon junction complex，**EJC**），它们在剪接的时候结合到外显子-外显子交界处。其中两个蛋白质 **Upf1** 和 **Upf2** 在哺乳动物 T 细胞有活性。如果用 RNAi 从细胞中去除其中任一

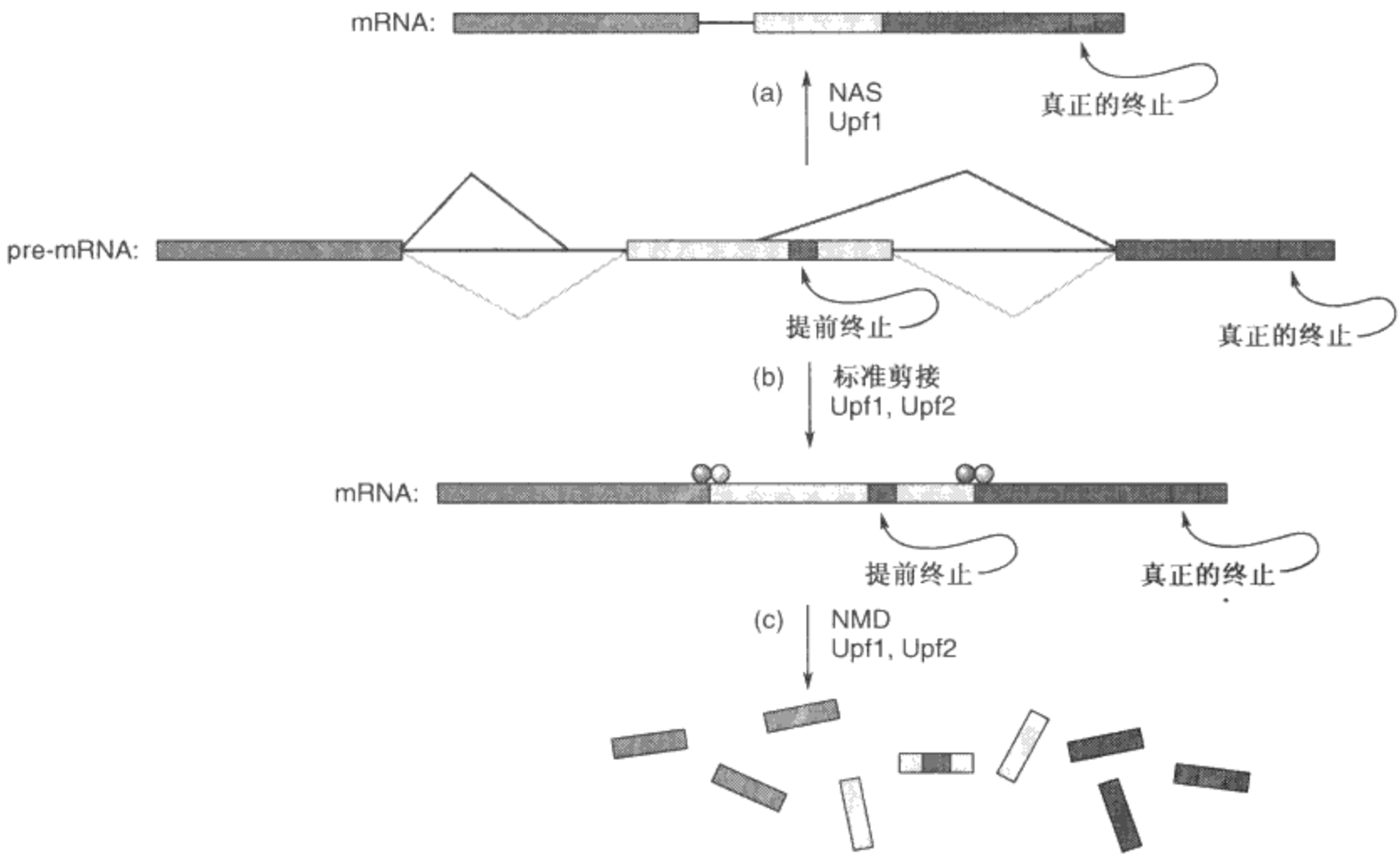


图 18.38　NAS 和 NMD 模型。（a）NAS。*Upf1*（也许与其他蛋白质联合）感知未来 mRNA 的可读框里有一个提前终止密码子，引入另一种选择性剪接模型（紫色）产生成熟 mRNA（顶部），该 mRNA 无提前终止密码子。（b）标准剪接（橙色）产生带有提前终止密码子的成熟 mRNA，*Upf1* 和 *Upf2* 结合在外显子/外显子交界处。（c）NMD。*Upf1* 和 *Upf2*（棕色、灰色小球）（也许与其他蛋白质联合）感知可读框内的提前终止密码子与第二外显子/外显子交界处的距离太近，引入 mRNA 的降解。

个（第 16 章），NMD 就会受到抑制。当这些蛋白质在终止密码子下游的足够远处与 mRNA 结合时，可将这个终止密码子识别为提前终止密码子，并被激活去降解这条 mRNA。另一方面，如果这些蛋白质离终止密码子相对较近，正在翻译 mRNA 的核糖体将去除之。

与刚才介绍的模型不同，Elisa Izaurralde 及同事在 2003 年报道，果蝇细胞中 NMD 不需要 EJC，由此提出 NMD 机制在不同生物中可能不同。此后，Allan Jacobson 及同事在 2004 年也报道了对酵母 NMD 的研究，表明其提前终止机制是自我中断。

具体讲，Jacobson 及同事利用趾纹分析（第 17 章）显示，一旦核糖体提前终止翻译，它们并不从 mRNA 解离，而是移动到起始密码子（AUG）的上游。这一行为可通过除去酵母 Upf1 蛋白，或在提前终止密码子附近放置一个正常 3′-UTR 而阻断。此外，含有提前终止密码子的 mRNA 可通过携带一个 poly（A）结合蛋白（Pab1p）而稳定下来。所有这些发现都支持这样一个模型，即核糖体通过靠近终止密码子的 3′-UTR 或 poly（A）的周围情况识别正常的终止密码子，并正常地终止。相反，核糖体通过提前终止密码子对这些正常线索的远离程度而准确识别提前终止密码子，并通过返回 AUG 的上游而异常终止翻译。理论上，任何真核细胞都应该能识别这种异常终止并降解有关的 mRNA，但是现在还不清楚 NMD 机制在真核生物有多大的一致性。

NAS 比 NMD 更神秘。当 NAS 机器检测到可读框内（不是可读框外）的提前终止密码子时，可调动剪接装置以一种不同方式剪接前体 mRNA，从成熟 mRNA 中除去提前终止密码子。但是这一机制引出一个有趣的问题：在前体 mRNA 被均匀剪接之前，NAS 机器是怎样检测未来的可读框的？

到目前为止，我们还没有该问题的答案，但是已经知道 Upf1 是 NAS 的一个重要因子，也是 NMD 的重要因子。Harry Dietz 及同事采用 RNAi 技术证明，Upf1 是 NAS 所必需的，而 Upf2 不是。随后他们细化了实验方法来回答是否 NMD 和 NAS 所需要的是 Upf1 的相同部分。研究采用了**等位基因特异性 RNAi**（allele-specific RNAi）技术：先构建一个经修饰的 *Upf1* 基因，使其不受双链 RNA 引起的 RNAi（阻止内源基因表达）的影响，然后将这个修饰的基因克隆到质粒上，转化细胞，该细胞里 RNAi 干扰内源 *Upf1* 基因的表达。修饰过的基因能恢复 NAS 和 NMD，否则它们会因 *Upf1* 表达丧失而被阻断。

接着，Dietz 及同事对修饰过 *Upf1* 基因的保守区域进行突变，其中一个突变敲除了修饰 *Upf1* 基因的恢复 NMD 能力，但不影响它对 NAS 的恢复能力。因此，虽然 NMD 和 NAS 都依赖 *Upf1*，但很显然它们依赖于该蛋白质的不同功能。

小结 真核生物通过两种不同的机制 NMD 和 NAS 来处理提前终止密码子。在哺乳动物细胞中，NMD 有一个下游不稳定元件，包括结合到 mRNA 的外显子-外显子交界处的 *Upf1* 和 *Upf2*。不稳定元件测定到终止密码子的距离，如果终止密码子位于足够远的上游区域，它可能就是提前终止密码子，激活下游的不稳定元件从而降解 mRNA。在酵母中，通过正常终止密码子附近的 3′-UTR 或 poly（A），细胞识别提前终止密码子，当核糖体停止在提前终止密码子处时，它移动到上游的 AUG 处，并标记待降解的 mRNA。NAS 机器能识别可读框中部的终止密码子，改变剪接模式，使提前终止密码子从成熟 mRNA 中剪接掉。与 NMD 一样，这一过程也需要 *Upf1*。

利用终止密码子插入非寻常氨基酸

大多数蛋白质只含有 20 个氨基酸如图 3.2 所示。但是，少数蛋白质需要非寻常氨基酸。最早发现的非寻常氨基酸如羟（基）脯氨酸（hydroxyprolin）是在由标准的 20 个氨基酸组成的蛋白质的翻译后修饰过程中产生的。最近已发现其他非寻常氨基酸如硒代半胱氨酸（selenocysteine）和吡咯赖氨酸（pyrrolysine）被直接加入到增长的多肽中。在这些例子中，有关的机制涉及利用编码区中间的终止密码子。细胞对这些终止密码子的翻译不是作为终止信号，而是作为非寻常氨基酸的密码子。

在蛋白质中发现的第一个非寻常氨基酸（“第 21 个氨基酸”）是硒代半胱氨酸，与半胱氨酸很像，只是在硫原子的位置是个硒原子。有些酶，如谷胱甘肽过氧化物酶和甲酸盐脱氢酶（formate dehydrogenase）缺乏硒代半胱氨酸时就没有活性，都需要一个硒代半胱氨酸残基作为其活性中心的一部分。那么这种非寻常氨基酸是怎样被加入蛋白质的呢？编码这些酶的基因在需要硒代半胱氨酸的位置含有 UGA 终止密码子。此外，缺乏硒时翻译就在这些终止密码子处提前停止。这些发现提示，细胞以某种方式将 UGA 密码子翻译为硒代半胱氨酸密码子，但是怎么实现的呢？

有一种特殊的 tRNA 带有识别 UGA 终止密码子的反密码子，可被正常的 seryl-tRNA 合成酶加载丝氨酸，然后在这个特殊 seryl-tRNA 中的丝氨酸被转换成硒代半胱氨酸，一种特殊的 EF-Tu 再将这个改变的氨酰-tRNA 送到核糖体上以响应在 mRNA 中

间的 UGA 密码子，不是编码区末端的 UGA 密码子。如果是后一种情况，则硒代半胱氨酸将响应真正的终止密码子而被加入，从而干扰终止。

因此，mRNA 内的 UGA 密码子只是召集 selenocysteinyl-tRNA 信号的一部分，mRNA 的其他部分也一定发挥作用了。例如，在甲酸脱氢酶 mRNA 中，内部 UGA 下游约 40nt 的区间，在另一个mRNA是在 3′-UTR 里，内部 UGA 下游约 1000nt 的区间。

“第 22 个氨基酸”是吡咯赖氨酸，其结构见图 18.39。与广泛分布的硒代半胱氨酸不同，吡咯赖氨酸至今只在某些产甲烷的古菌中发现。另一个不同是，硒代半胱氨酸是由 seryl-tRNA 上的正常氨基酸（丝氨酸）产生的，而吡咯赖氨酸是先合成然后再通过特殊的 pyrrolysyl-tRNA 合成酶加载到特殊的 tRNA 上的。这是所发现的第 21 种氨酰-tRNA 合成酶，唯一一个不同于其他 20 种正常氨基酸加载正常 tRNA 的氨酰-tRNA 合成酶。

*E. coli*细胞不能正常地将吡咯赖氨酸加入其蛋白质中，但是 Joseph Krzycki 及同事在 2004 年发现，如果对细胞添加 3 种东西，就可以赋予*E. coli*细胞这种能力。一个是编码特殊 tRNA 的基因，另一个是编码特殊的 pyrrolysyl-tRNA 合成酶的基因，最后是吡咯赖氨酸自身。此外，他们发现 tRNA 能够接受体外预形成的吡咯赖氨酸，明显提示这就是它在体内作用的方式。

与硒代半胱氨酸一样，吡咯赖氨酸响应终止密码子而被加入增长的多肽中，但这种终止密码子是 UAG 而不是 UGA。由此提示，这个特殊 tRNA 的反密码子是 5′-CUA-3′，情况确实如此。

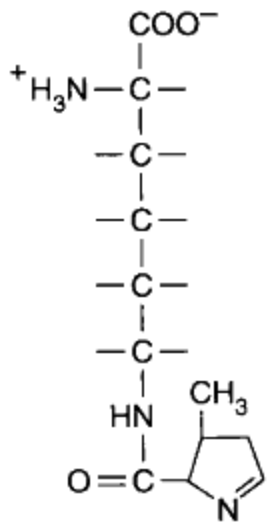


图 18.39　吡咯赖氨酸。

小结　非寻常氨基酸硒代半胱氨酸和吡咯赖氨酸分别响应终止密码子 UGA 和 UAG 而被加入到增长的多肽中，机制如下：①硒代半胱氨酸：特殊的 tRNA 具有识别 UGA 的反密码子被丝氨酸加载，丝氨酸再被转换成硒代半胱氨酸，selenocysteyl-tRNA 被特殊的 EF-Tu 护送到核糖体上；②吡咯赖氨酸：特殊的 pyrrolysyl-tRNA 合成酶将预先形成的吡咯赖氨酸连接到特殊的 tRNA 上，该 tRNA 的反密码子识别密码子 UAG。

18.5　翻译后

翻译的故事并不是以终止结束的。蛋白质必须正确地折叠，而且核糖体需要从 mRNA 释放以便参与后面的翻译。严格地讲，折叠过程不是在翻译后发生的，而是新生多肽合成中发生的共翻译事件。然而单独处理可能更方便一些，因为它与刚讨论过的起始、延伸、终止事件都没有直接的关系。我们先来讨论折叠问题，然后讨论核糖体释放问题。

新生蛋白质的折叠

天然蛋白质都是折叠的，以便疏水区埋在蛋白质的内部，避开细胞的水环境。但是多数蛋白质自身并不能正确折叠，就像被热激去折叠的蛋白质一样，需要分子伴侣的帮助（第 8 章）。问题是新生多肽任何暴露的疏水部分都试图找到能与它相互作用的疏水部分，以便躲开围绕它们的水分子。但是最近的疏水区往往是错误的搭档，这种相互作用会导致错误的折叠，进而使蛋白质失活。实际上，有些错误折叠的蛋白质，比如参与牛海绵组织脑病（BSE，或“疯牛病”）的一种蛋白质对细胞就是剧毒的。

因此，细胞需要一种机制来隐藏新生多肽的疏水部分，直到正确的搭档产生出来。分子伴侣可以完成这一任务，它们将暴露的疏水蛋白区装进自己的疏水口袋中，防止与其他暴露的疏水区的不适当结合。但是*E. coli*有一个特殊的分子伴侣，叫做**触发因子**（trigger factor），可以与核糖体大亚基结合，用一个疏水筐盛装新合成的疏水区，保护它们避开水环境。

要想了解触发因子是如何工作的，最好能有分子伴侣结合到其核糖体泊位的晶体结构，但问题是已经结晶出来的唯一的大核糖体亚基是古菌（*Haloarcula marismortui*）的（第 19 章），而古菌没有触发因子。所以，Nenad Ban 及同事结晶出完整的*E. coli*触发因子以便观察其形状，然后与古菌大核糖体亚基一起结晶出*E. coli*触发因子的核糖体结合部位，期望古菌与细菌的核糖体结合位点非常保守，可以形成跨界的复合物。

庆幸的是，这个策略是有效的！触发因子的结合位点（在核糖体蛋白 L23 上）在古菌和细菌中高度保守，古菌的核糖体亚基可以结合细菌触发因子。

触发因子单独的晶体结构是一个“卧龙”形的，具有头、背、臂、尾，如图 18.40 所示。基于 50S 核糖体亚基与触发因子尾部结构域的共结晶结构，Ban 及同事按图 18.40 所示放置触发因子，让“卧龙”底朝天，这样尾和臂的疏水表面处于完美的位置，刚好接住从核糖体出口通道出来的新生多肽。这就有效地隐藏了新生多肽任何暴露的疏水区，直到它们与合适搭档的疏水区结合。

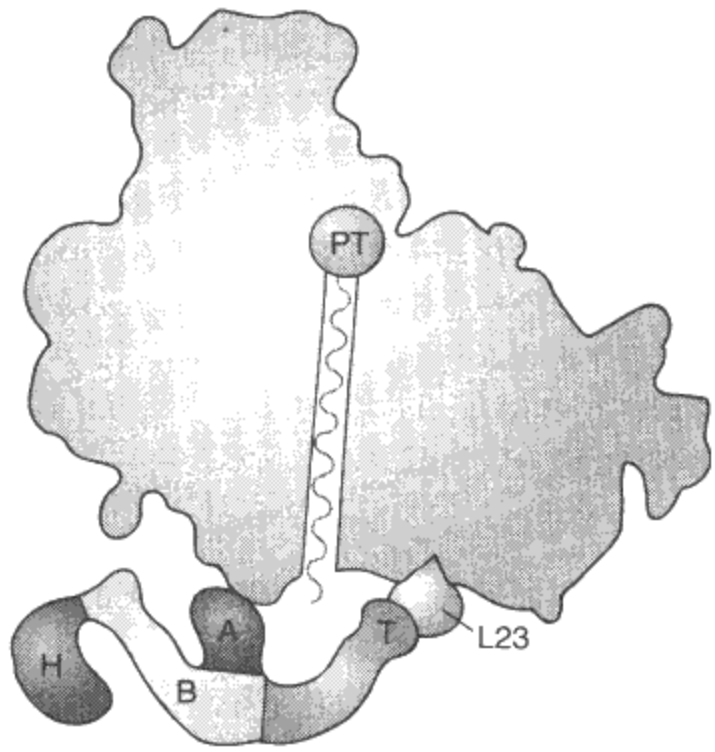


图 18.40　触发因子结合核糖体的模型。伴侣蛋白触发因子像一条底朝天的卧龙结合在核糖体的下部盖住出口通道。在这个位置，触发因子的疏水结构域［臂（A）、尾（T）］可以在新生多肽从出口通道出来时接住它们的疏水区，使它们处于疏水环境直到能与该新生多肽的其他疏水区结合，从而促进正确折叠。触发因子的其他结构域是头（H）和背（B）。L23 是大核糖体亚基的一个蛋白质，是触发因子的主要结合位点。PT 是在通道出口处肽酰转移酶的位点。（*Source*：Adapted from Ferbitz，L.，T. Maier，H. Patzelt，B. Bukau，E. Deverling，and N. Ban，Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins，*Nature* 431：593，2004.）

触发因子对*E. coli*并不重要，因为细菌有 DnaK 分子伴侣备用系统。这是一种独立自由的蛋白质，不像触发因子这种核糖体结合蛋白。它不是以疏水筐接住新生蛋白质，而是有一个疏水的拱门来保护新生蛋白质暴露的疏水区，直到正确折叠。古菌和真核生物完全缺乏类触发因子蛋白质，所以它们的新生蛋白质的正确折叠完全依赖于独立自由的分子伴侣。

小结　多数新生多肽自身不能正确折叠，而是需要分子伴侣的帮助。*E. coli*细胞有一种叫做触发因子的蛋白质可以与核糖体结合使新生多肽从核糖体的出口通道一出来就被接住，这样新生多肽的疏水区就被保护起来避免产生不当结合，直到有合适的搭档出现。古菌和真核生物没有触发因子，它们必须利用游离的分子伴侣，这在细菌中也存在。

核糖体从 mRNA 的释放

对于终止的早期研究采用的模型系统只包括 AUG 和 UAG 作为 mRNA 的类似物，这些研究中没有检测对核糖体释放的需求，部分原因是这些模型的 mRNA 可自发地从核糖体上解离。

后来 A. Kaji 及同事发现了一个能将核糖体从正常 mRNA 释放的蛋白质因子，命名为**核糖体循环因子**（ribosomal recycling factor，**RRF**）。1994 年，他们证明 RRF 对细菌生长至关重要。RRF 基因的温度敏感突变体在，生长的延滞期转入非允许温度时会导致细菌死亡，在生长的对数期转入非允许温度，会导致生长停止。因此，翻译终止后核糖体从 mRNA 的释放很重要。

Kaji 及同事从细菌海栖热袍菌（*Thermotoga maritima*）中纯化出 RRF，并用下列分析进行了检测。他们用嘌呤霉素处理细菌的多核糖体，释放新生多肽，使每个核糖体都剩下两个去氨酰-tRNA，一个在 P 位点，另一个在 E 位点。因此，这些多核糖体中的每个核糖体都类似于一个刚刚经过翻译终止的核糖体，除了在 A 位点没有终止密码子。然后，对这些嘌呤霉素处理的核糖体加入 RRF，使多核糖体变成单体。纯化后，Kaji 小组与 Anders Liljas 小组合作，测定了 RRF 蛋白的晶体结构。

RRF 的晶体结构是惊人的——一个几乎完美的 tRNA 仿制品！图 18.41 是 *T. maritima* 的 RRF 与 $tRNA^{Phe}$ 结构的叠加图。它们几乎完美地重合，RRF 唯一缺少的是填充由 tRNA 末端 CCA 正常占据的位置的氨基酸，以及一小段反密码子。基于该结构和其他信息，Kaji 及同事认为 RRF 就像氨酰-tRNA 一样与 A 位点结合，因而在 EF-G 存在时允许发生移位，然后以某种方式将核糖体从 mRNA 上释放。

2002 年，Kaji 小组与 Noller 小组合作，用**羟基自由基探测**（hydroxyl radical probing）对 RRF-核糖体复合物的结构进行了研究。他们采用的方法如下：首先，利用定点突变用丝氨酸取代 RRF 分子中唯一的半胱氨酸。然后，突变这个仍保持活性的无半胱氨酸 RRF，在贯穿该分子的 10 个不同位点各置入一个半胱氨酸。每个带有单一半胱氨酸的 RRF 分子都能被耦合到带有 Fe^{2+} 的分子上。然后，RRF-Fe^{2+} 能被结

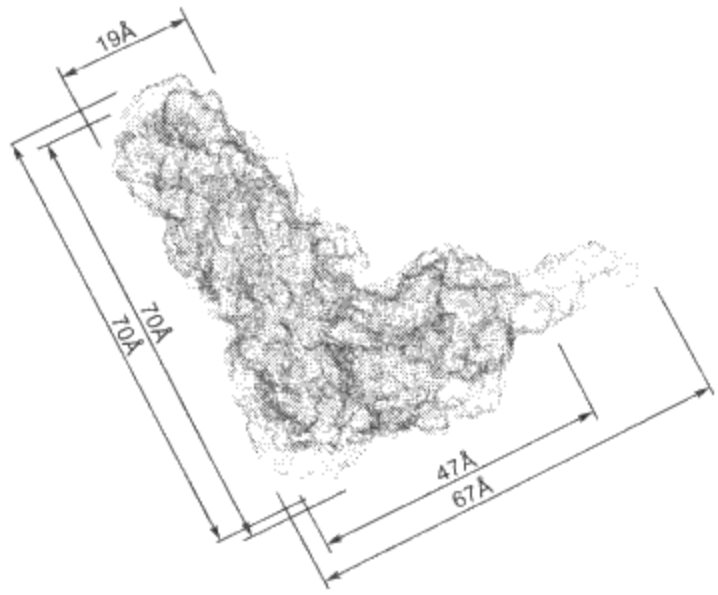


图 18.41　RRF 和 tRNA 的结构叠加图。海栖热袍菌 RRF（蓝色）和酵母 $tRNA^{Phe}$（红色）的表面被叠加在一起显示出它们有极大的相似性。（*Source*：From Kaji et al.，*Cold Spring Harbor Symposia* 66：p. 518. © 2002.）

合到核糖体上。Fe^{2+} 产生羟基打断附近的 rRNA 片段，打断情况可通过引物延伸法进行检测（第 5 章）。因为我们已准确地知道 16S 和 23S rRNA 的各部分在核糖体上的定位（第 19 章），可以在核糖体具体位置上定位出 RRF 各部分的图谱。

这个实验证明，尽管 RRF 与 tRNA 在结构上近乎完美地相似，但它在与核糖体结合方面与 tRNA 并不相同。它与核糖体 A 位点结合的方向完全不同于 tRNA 对 A 位点的结合（图 18.42a）。这一结果对 Kaji 及同事的简单模型提出了疑问。事实上，它还产生了 RRF 是如何以这种方式结合到核糖体上的问题，因为 RRF 的末端会与结合在 P 位点的去氨酰 tRNA 的受体臂重叠。但是，Kaji 和 Noller 及同事注意到被嘌呤霉素或可能是 RF1 或 RF2 去氨酰的 tRNA，并不以纯 P 位点结合状态存在，而是如 Noller 及同事所证明的，处于 P/E 杂合状态，其受体臂末端在 E 位点，反密码子在 P 位点。以这种位置，它不会干扰 RRF 与核糖体的结合，如图 18.42b 所示。

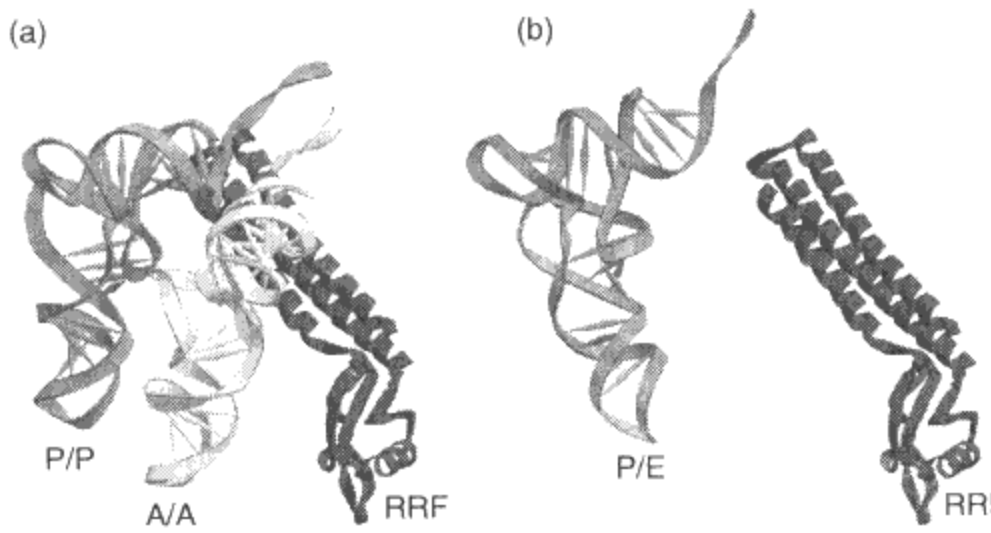


图 18.42　RRF 在核糖体上的位置模型。（a）RRF（红色）相对于在纯 A 位（A/A，黄色）和 P 位（P/P，橙色）tRNA 的位置。（b）RRF（红色）相对于结合在杂合的 P/E 位 tRNA 的位置（橙色）。（*Source*：Reprinted from Cell v. III，Lancaster et al.，p. 444 © 2002，with permission from Elsevier Science.）

那么，RRF 结合到 A 位点后会发生什么呢？这仍然很不清楚，尽管我们知道它与 EF-G 作用，促进核糖体从 mRNA 上释放。有时，它只释放 50S 亚基，留下 30S 亚基由另外的机制如通过与 IF3 结合来释放。在这种情况下，由于核糖体不能被完整释放，就不需要 IF1 和 IF3 合作来解离两个亚基了。

小结　翻译终止后，核糖体不能自发地从 mRNA上解离下来，需要核糖体循环因子（RRF）和 EF-G 的帮助。RRF 与 tRNA 的结构十分类似，能结合在核糖体的 A 位点，但并不是通常 tRNA 结合的位置。然后，它与 EF-G 合作释放 50S 核糖体亚基，或整个核糖体。释放机制现在还不清楚。

总结

信使 RNA 是以 5′→3′方向阅读的，与其合成的方向相同。蛋白质是以氨基端到羧基端方向合成的，表明氨基末端的氨基酸是先加上去的。

遗传密码是 mRNA 上的一套 3 碱基密码，指导核糖体将特定的氨基酸加入多肽。这种密码是非重叠性的，即每个碱基只是一个密码子的一部分。该密码也缺乏间隔或间隔符，即 mRNA 编码区的每个碱基是一个密码子的一部分。共有 64 个密码子，3 个为终止信号，其他密码子都编码氨基酸，这意味着密码是高度简并的。

遗传密码的简并性部分是由同工 tRNA 提供的，同工 tRNA 结合同一个氨基酸但识别不同的密码子，其余部分由摇摆来决定，其中密码子第三位碱基被允许稍微偏离其正常位置，与反密码子形成非 Watson-Crick 配对，因而同一个氨酰-tRNA 可以与一个以上的密码子配对。摇摆碱基对包括 G-U（或 I-U）和 I-A。

遗传密码并不是严格通用的，在某些真核生物细胞核和线粒体基因组中，以及在至少一种细菌中，

标准遗传密码的终止密码子可以编码氨基酸，如色氨酸和谷氨酸。在几种线粒体基因组及至少一种酵母核基因组中，密码子的含义可从一种氨基酸变到另一种氨基酸。这些变异的密码子与它们可能基于此变来的标准密码子有着紧密的联系。

肽链延伸以下列 3 个步骤重复发生：①EF-Tu 及 GTP 结合氨酰-tRNA 到核糖体的 A 位点；②肽酰转移酶催化 P 位点的肽与 A 位点新到来的氨酰-tRNA形成肽键，使肽链延长一个氨基酸，并将其移到 A 位点；③延伸因子 EF-G 以及 GTP 使延伸的肽酰-tRNA 及其 mRNA 密码子移位到 P 位点。

嘌呤霉素的化学结构与氨酰-tRNA 类似，可结合到 A 位点并偶联 P 位点的肽，将其以肽酰-嘌呤霉素形式释放。此外，如果在 A 位点有肽酰-tRNA，嘌呤霉素不能与核糖体结合，肽段也就得不到释放。由此定义了核糖体的两个位点：嘌呤霉素反应位点（P 位）和嘌呤霉素非反应位点（A 位）。在 70S 起始复合物上的 fMet-tRNA$_f^{Met}$ 是嘌呤霉素反应性的，故位于 P 位点。核糖体结合实验和结构研究鉴定出第三个位点（E 位点），由去氨酰-tRNA 结合这类 tRNA 在离开核糖体时，与 E 位点结合。

由 EF-Tu、氨酰-tRNA 和 GTP 形成的三元复合物可将氨酰-tRNA 转移到核糖体的 A 位点，此过程不需要 GTP 水解。随后，EF-Tu 以其核糖体依赖性 GTP 酶活性水解 GTP，形成 EF-Tu-GDP 复合物并从核糖体解离。EF-Ts 通过用 GTP 交换与 EF-Tu 结合的 GDP，使 EF-Tu-GTP 复合物再生，氨酰-tRNA的加入又重新构成三元复合物进入新一轮肽链延伸。

蛋白质合成机器在延伸过程中通过一个两步过程来保证翻译的准确性。首先，在 GTP 水解之前清除携带错误氨酰-tRNA 的三元复合物。如果这个监控失败，仍能通过校正，在错误氨基酸加入延伸链之前清除错误的氨酰-tRNA。根据 Thompson 的假说，这两个校正过程依赖于不正确密码子与反密码子间碱基的弱配对作用，从而确保解离作用比 GTP 水解或肽键的形成更快发生。翻译速度和准确性之间的平衡是精密的，如果肽键形成过快，不正确氨酰-tRNA 不能及时离开核糖体，它们的氨基酸就会被整合到蛋白质，如果肽键形成速度过慢，蛋白质的合成又不能满足生物体成功生长的需要。

叫做肽基转移酶的核糖体酶催化合成肽键，其活性位于 50S 核糖体亚基上，23S rRNA 含有肽基转移酶的活性中心。

每一次移位使 mRNA 通过核糖体移动一个密码子的长度（3nt）。尽管移位活性可以在体外没有 EF-G 和 GTP 时表现出来，但是 GTP 和 EF-G 是移位所需要的。EF-G 必须依靠 GTP 水解释放的能量从核糖体上释放出来，开始新一轮延伸。EF-Tu-tRNA-GDPNP 三元复合物和 EF-G-GDP 二元复合物的三维结构已经由 X 射线晶体图分析确定，正如所预计的那样，它们非常相似。

mRNA 分子中的琥珀、赭石和卵白石突变分别产生 UAG、UAA 和 UGA 3 个终止密码子，造成肽链翻译的提前终止。这 3 个密码子也是 mRNA 编码区末端的正常终止信号。大部分抑制子 tRNA 具有改变的反密码子，能够识别终止密码子，通过插入一个氨基酸允许核糖体通过终止密码子移动到下一个密码子而阻止终止。

原核生物翻译终止受 3 种因子 RF1、RF2 和 RF3 的介导。RF1 识别终止密码子 UAA 和 UAG，RF2 识别 UAA 和 UGA。RF3 是一个 GTP 结合蛋白，促进 RF1 和 RF2 与核糖体结合。真核细胞有两个释放因子 eRF1 和 eRF3。eRF1 识别所有 3 个终止密码子，eRF3 是一种核糖体依赖性 GTP 酶，能帮助 eRF1 释放翻译成熟的肽链。

原核生物通过 tmRNA 介导的核糖体恢复来处理非终止 mRNA。与丙氨酰-tRNA 类似的丙氨酰-tmRNA，结合到停滞在非终止 mRNA 上的核糖体的空 A 位上，将其丙氨酸添加到停滞的多肽上。然后核糖体转而翻译 tmRNA 的 ORF，在终止前给多肽再添加 9 个氨基酸。这些额外的氨基酸标定待降解的多肽，核酸酶降解非终止 mRNA。真核生物在非终止 mRNA poly（A）尾末端的核糖体将 Ski7p-外来体复合物吸引到空 A 位，然后 Ski 复合物结合到 A 位，外来体正好定位于非终止 mRNA 的末端，降解 RNA。估计异常多肽同时也被降解了。

真核生物通过两种不同的机制（NMD 和 NAS）来处理提前终止密码子。在哺乳动物细胞中，NMD 有一个下游不稳定元件，包括结合到 mRNA 的外显子-外显子交界处的 Upf1 和 Upf2，不稳定元件测定与终止密码子的距离。如果终止密码子位于足够远的上游区域，它可能就是提前终止密码子，激活下游的不稳定元件从而降解 mRNA。在酵母中，终止密码子附近缺少正常 3′-UTR 或 poly（A）会被识别为非正常终止密码子。NAS 机器能识别可读框中部的终止密码子，改变剪接模式，使提前终止密码子从成熟 mRNA 中剪接掉。与 NMD 一样，这一过程也需要 Upf1。

非寻常氨基酸硒代半胱氨酸和吡咯赖氨酸分别响应终止密码子 UGA 和 UAG 被加入增长的多肽中，机制如下：①硒代半胱氨酸：特殊的 tRNA（具有识别 UGA 的反密码子）被丝氨酸加载，丝氨

酸再被转换成硒代半胱氨酸，selenocysteyl-tRNA 被特殊的 EF-Tu 护送到核糖体上；②吡咯赖氨酸：特殊的 pyrrolysyl-tRNA 合成酶将预先形成的吡咯赖氨酸 连接到特殊的 tRNA 上，该 tRNA 的反密码子识别密码子 UGA。

多数新生多肽自身不能正确折叠，需要分子伴侣的帮助。*E. coli*细胞有一种叫做触发因子的蛋白质可以与核糖体一起使新生多肽从核糖体的出口通道一出来就被接住，这样新生多肽的疏水区就被保护起来而避免不当结合，直到有合适的搭档。古菌和真核生物没有触发因子，它们必须利用游离的分子伴侣，这在细菌中也存在。

翻译终止后，核糖体不能自发地从 mRNA 上解离，需要核糖体循环因子（RRF）和 EF-G 的帮助。RRF 与 tRNA 结构十分类似，能结合在核糖体的 A 位点，但并不是通常的 tRNA 所结合的位置。然后与 EF-G 合作释放 50S 核糖体亚基，或整个核糖体。释放机制现在还不清楚。

复习题

1. 叙述实验并给出结果，说明翻译是从蛋白质的氨基末端开始的。

2. 怎样知道 mRNA 的阅读是 5′→3′方向？

3. 怎样知道遗传密码是：（a）非重叠的；（b）无间隔的；（c）三联体的；（d）简并的？

4. 叙述实验并给出结果，揭示两个密码子对应于一个氨基酸。

5. 图示一个摆动碱基对，不必显示所有原子的位置，只需要碱基对的形状，并与 Watson-Crick 碱基对的形状对比。摆动在翻译中的重要性是什么？

6. 图示原核生物的翻译延伸过程。

7. 图示噁唑霉素的作用模型。

8. 叙述实验并给出结果，说明 fMet-tRNA$_f^{met}$ 占据了核糖体的 P 位点。

9. 叙述实验并给出结果，说明 EF-Ts 从 EF-Tu 上释放 GDP。

10. 氯霉素阻止翻译的什么步骤？

11. 图示 EF-Tu 和 EF-Ts 在翻译中的作用。

12. 提供 EF-Tu、GTP、氨酰-tRNA 形成三元复合体的证据。

13. 叙述实验并给出结果，说明核糖体 RNA 可能是肽基转移酶的催化因子。

14. 蛋白质合成中什么是初始识别和校正？

15. 叙述实验并给出结果，说明在移位中 mRNA以 3nt 单位移动。

16. 叙述实验并给出结果，说明 EF-G 和 GTP 都是移位所必需的。

17. 描述实验，说明 GTP 水解先于移位。

18. 提供琥珀密码子是翻译的终止密码子的直接证据。

19. 提供琥珀密码子是 UAG 的证据。

20. 解释琥珀抑制子是如何工作的？

21. 提供琥珀抑制子是 tRNA 的证据。

22. 描述分析释放因子的实验。

23. RF1、RF2 和 RF3 的作用各是什么？

24. 怎样知道在 RF1 和 RF2 中哪个氨基酸识别终止密码子？

25. eRF1 和 eRF3 的作用是什么？

26. 图示原核生物处理非终止 mRNA 的机制。

27. tmRNA 与 tRNA 之间的何种差异限制了 tmRNA 紧密结合核糖体的能力？细胞如何处理这些缺陷？

28. 图示哺乳动物细胞处理非终止 mRNA 的机制。

29. 图示真核生物处理提前终止密码子的两种机制。

分析题

1. 以下作用对 G 蛋白活性的影响效果是什么？

a. 抑制其 GAP；

b. 抑制其鸟嘌呤核苷酸交换蛋白质。

2. 你分离出了一株*E. coli*突变体，其氨酰-tRNA 合成酶在温度升至 42℃时使反密码子为 3′-UUC-5′的 tRNA 被加载天冬氨酸。你预期这对细胞在 42℃的蛋白质合成有什么影响，为什么？然后分离出另一个突变体，它能抑制第一个突变。你追踪第二个突变到 tRNA 基因，你认为第二个突变体中哪个 tRNA 将被改变，在哪儿改变？推测这种改变的性质。

3. 考虑短 mRNA：5′-AUGGCAGUGCCA-3′，先假设密码完全重叠，再假设非重叠，回答下列问题：

a. 在这个寡聚核苷酸中会有多少密码子？

b. 如果第二个 G 变成 C，有多少密码子将被改变？

4. 如果是下列改变，对可读框和基因功能会产生什么影响？

a. 在 mRNA 中间插入 2 个碱基；

b. 在 mRNA 中间插入 3 个碱基；

c. 在一个密码子中插入 1 个碱基并从下一个密码子中减去一个碱基。

5. 如果密码子有 6 碱基长，你认为从重复的四核苷酸［如 poly（UUCG）］会产生什么样的产物？

6. 如果密码子是 4 碱基长，在遗传密码中会有

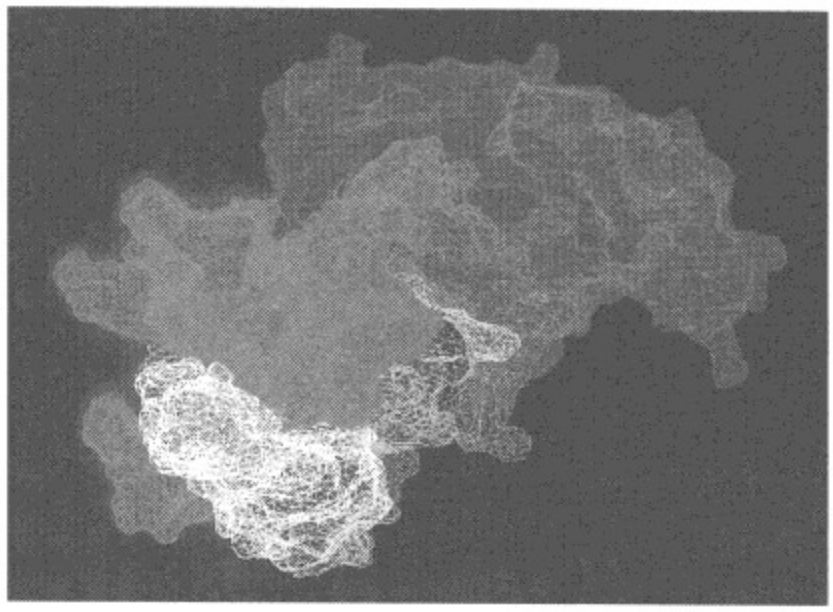
多少密码子？

7. 某个赭石抑制子响应赭石密码子插入谷氨酰胺，产生这种抑制子菌株的 $tRNA^{Gln}$ 反密码子的最可能改变是什么？

8. 描述使有机体获得将 pyrrolysine 加入其蛋白质的能力的进化变化，你认为这些变化以什么顺序进行？为什么？提示：见 Wang L. (2003). Expanding the genetic code. Science 302：584～585.

翻译　马　纪　校译　张富春

第 19 章　核糖体和转运 RNA



结合到氨酰-tRNA 合成酶上的转运 tRNA 的计算机生成模型 *Copyright © Kenneth Eward / BioGraft / SS / Photo Researchers, Inc.*

在第 3 章我们学习了蛋白质翻译的几方面内容，知道核糖体是细胞的蛋白质工厂，转运 RNA 起着至关重要的适配器作用，其一端结合着氨基酸，另一端连接着 mRNA 密码子。第 17、18 两章进一步讨论了翻译的起始、延伸和终止机制，但没有深入讨论核糖体和 tRNA。本章将继续讨论翻译，更详细地研究这两个重要成分。

19.1　核糖体

第 3 章介绍了*E. coli*核糖体是沉降系数为 70S 的二元（two-part）结构，该结构的两个亚基是 **30S** 和 **50S** 核糖体亚基。在第 3 章我们也了解到小亚基的功能是解码 mRNA，而大亚基通过肽键将氨基酸连接在一起。在本节我们将重点讨论细菌核糖体的整体结构、组成、装配和功能。

70S 核糖体的精细结构

X 射线结晶学可提供最好的结构信息，但对于像核糖体这样大的不对称物体则十分困难。尽管如此，Harry Noller 及同事在获得适于 X 射线晶体学研究的嗜热菌的核糖体晶体方面还是取得了成功。到 1999 年，他们已经获得了这些核糖体的晶体结构，其研究在当时提供了完整核糖体的最详细的结构信息，分辨率高达 7.8Å。

随后，Noller 及同事在 2001 年结晶出嗜热菌的 70S 核糖体复合物，外加 mRNA 类似物以及结合到核糖体 P 和 E 位点的 tRNA。由这些晶体获得分辨率为 5.5Å 的结构，较之以前的结构有相当大的提高。研究者还获得相同复合物的 tRNA 结合或未结合 A 位点的晶体，并且通过比较获得了 tRNA 在 A 位点的结构，其分辨率达到 7Å。

图 19.1 显示 70S 核糖体的晶体结构。图 19.1a～d 从前面、右侧、背面、左侧 4 个方向显示核糖体。各部分的颜色表示分别是：30S 亚基的 16S rRNA 为蓝绿色、30S 蛋白为蓝色；50S 亚基的 23S rRNA 为灰色、5S rRNA 为深蓝色、50S 蛋白为紫色。在 A、P、E 位点的 tRNA 分别用金色、橙色、红色表示，由于它们位于核糖体两个亚基间的裂缝内，在图 19.1a～d 上很难看到。大部分核糖体蛋白都已被鉴定出，注意 L9 蛋白从核糖体主体的侧边伸出很远（图 19.1a 的左侧）。图 19.1e 显示核糖体的俯视图，图 19.1 中 3 个 tRNA 清晰可见，注意 3 个 tRNA 的反密码子茎都向下插入位于底部的 30S 亚单位。

图 19.1f 和图 19.1g 显示分离的两个核糖体亚基，以便揭示 tRNA 结合的位置。其中 30S 亚基已沿垂直轴心旋转了 180°，所以能看到 3 个 tRNA。注意 tRNA 结合的裂缝在两个亚基中都与 rRNA 排成一行。在这个视角，蛋白质大多在外围。这一发现提示，是 rRNA 而不是蛋白质主宰着在 30S 亚基解码和 50S 亚基多肽合成中至关重要的与 tRNA 的相互作用。此外，核糖体与 3 个 tRNA 保守部分的相互作用使其能以完全相同的方式结合所遇到的不同 tRNA。

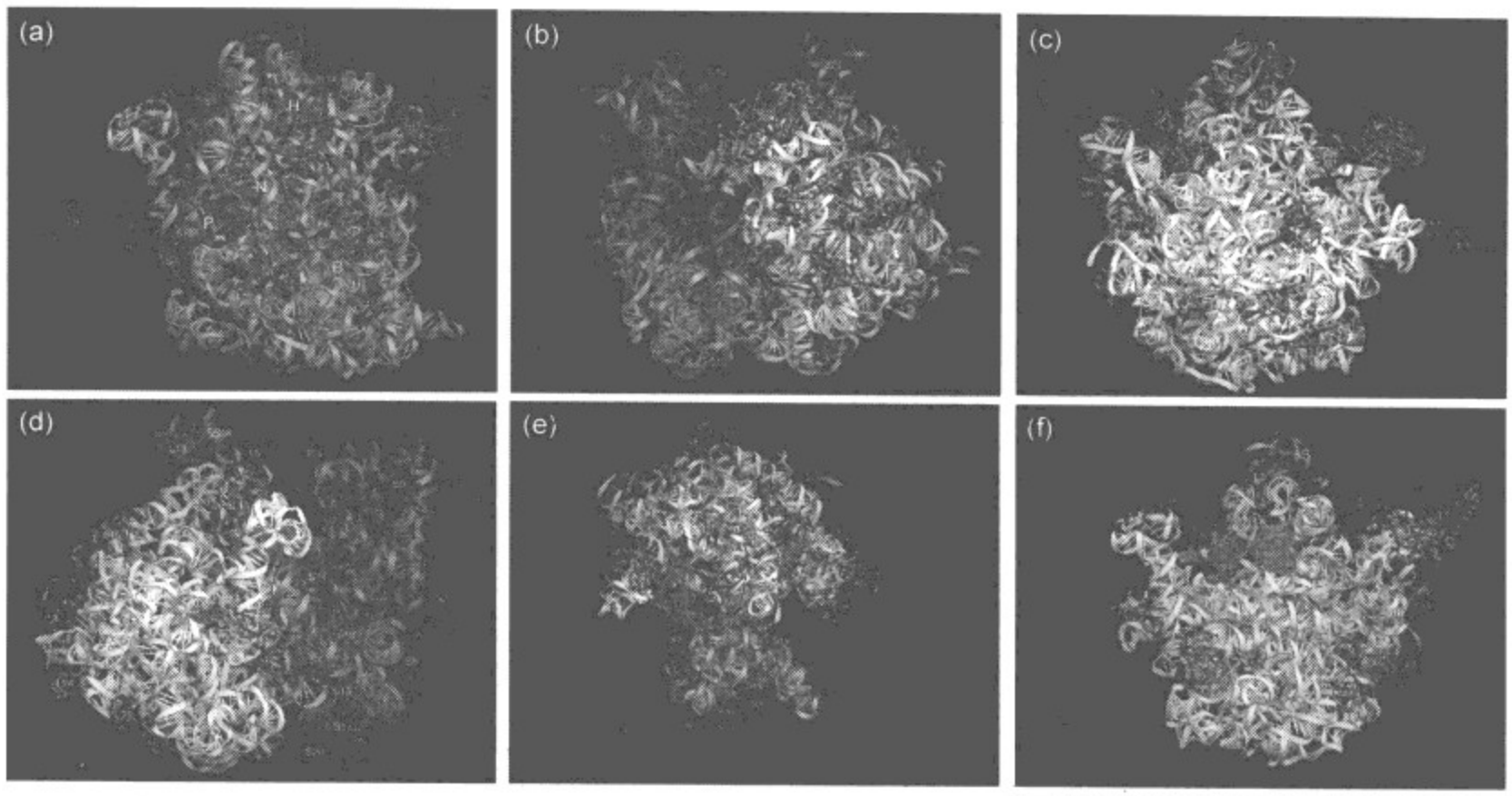


图 19.1　嗜热菌 70S 核糖体的晶体结构。（a～d）从 4 个不同方向展示核糖体，每个核糖体相对于前一个都沿垂直轴旋转 90°。在（a）中，30S 亚基在 50S 亚基前面。代表颜色：16S rRNA，青色；30S 核糖体蛋白，蓝色；23S rRNA，灰色；5S rRNA，深蓝；50S 核糖体蛋白，紫色；在 A、P、E 位点的 tRNA 分别是金色、橙色和红色；核糖体蛋白用数字标明。（e）俯视图：50S 亚基在顶部，30S 亚基在底部，3 个 tRNA 在中间。（f）和（g）分别为 50S 和 30S 亚基的作用面，30S 亚基旋转了 180°以揭示作用面上的 tRNA。（*Source*：From Yusupov et al.，*Science* 292：p. 885. © 2001 by the AAAS.）

还需注意，图 19.1g 中，反密码子茎环向下插入 30S 亚单位。位于 A 位和 P 位 tRNA 的反密码子在 10Å 以内相互接近，但这一距离还不足以使它们与邻近的密码子结合，核糖体通过将 A 位和 P 位密码子间的 mRNA 扭转 45°解决了这一问题（图 19.2），由此使两个密码子以合适的方向被 tRNA 解码。图 19.1（f）显示在 50S 亚基中位于 A 位和 P 位的tRNA彼此紧密地靠近，虽然在这个视角难以看到，但这两个 tRNA 的受体茎都插入 50S 亚基的肽基转移酶口袋内，在 5Å 距离内彼此靠近。这一紧密靠近是必要的，因为结合于这两个 tRNA 上的氨基酸和多肽在肽键形成过程中必须结合在一起。

70S 核糖体晶体结构揭示了亚基间的 12 处连接（亚基间桥，intersubunit bridge），见图 19.3。多数亚基间桥由 RNA 而不是蛋白质组成，实际上所有 tRNA-结合位点附近的亚基间桥都只有 RNA。注意，桥 B2a、B3、B5、B6 都只有一个 30S 亚基的 16S rRNA 螺旋域（螺旋 44）（图 19.3）。该螺旋是连接两个亚基的主要结构，而且我们将在本章后面看到它在密码子-反密码子识别中也发挥作用。由于 tRNA 从 A 位移动到 P 位再到 E 位需要移动 20～50 Å 的距离，所以至少其中一些亚基间桥是动态的断开和再形成，从而使移位能够发生。

图 19.4 是更模式化的核糖体图，突出 3 个重点：第一，在两个核糖体亚基间存在一个大洞，可容纳 3 个 tRNA；第二，tRNA 通过其反密码子末端与 30S 亚基相互作用，同时也结合与 30S 亚基结合的 mRNA；第三，tRNA 通过其受体茎与 50S 亚基相互作用，这是合理的，因为肽基转移酶反应期间受体茎必须在一起，而这一过程是在 50S 亚基上发生的，P 位点的肽酰-tRNA 受体茎所连的肽与 A 位点氨酰-tRNA 受体茎所连的氨基酸相连接。

2005 年，Jamie Doudna Cate 及同事取得了一个重大突破，他们得到了分辨率为 3.5Å 的*E. coli* 70S 核糖体晶体结构！这一突破不仅在于获得了目前所得到的 70S 核糖体的最高分辨率，更在于它是人们探寻已久的*E. coli* 70S 核糖体结构，生物化学和遗传学资料已经对它的结构补充了几十年了。在获得这一

图 19.2　在 A 位和 P 位密码子-反密码子碱基配对的立体图。3 个位点的 tRNA 用不同颜色表示（A，金黄色；P，橙色；E，红色）。密码子和反密码子的碱基如图底部的棒型所示。注意密码子间 mRNA 的 45°扭曲。未显示 E 位点 tRNA 的反密码子，因为它不与 mRNA 碱基配对。（*Source*：From Yusupov et al.，*Science* 292：p. 893. © 2001 by the AAAS.）

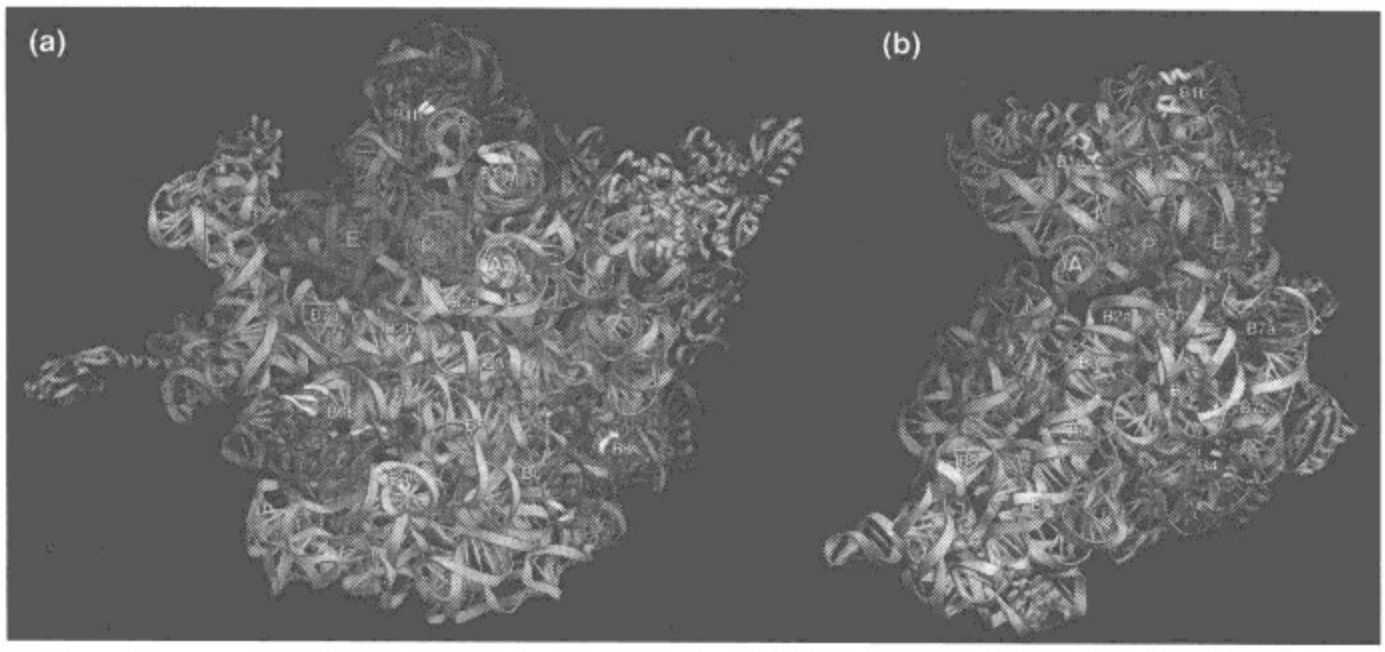


图 19.3　亚基间桥界面示意图。（a）和（b）分别是 50S 和 30S 亚基。颜色表示：在两个亚基中大 rRNA 为灰色、5S rRNA 为深蓝色、蛋白质为淡蓝色。tRNA 的颜色同图 19.1，分别是金色、橙色和红色，亚基间的 RNA-RNA 桥以粉色表示，蛋白质-蛋白质桥以黄色表示。所有桥都以数字编号（B1a、B1b、B2a 等）。（*Source*：From Yusupov et al.，Science 292：p. 890. © 2001 by the AAAS.）

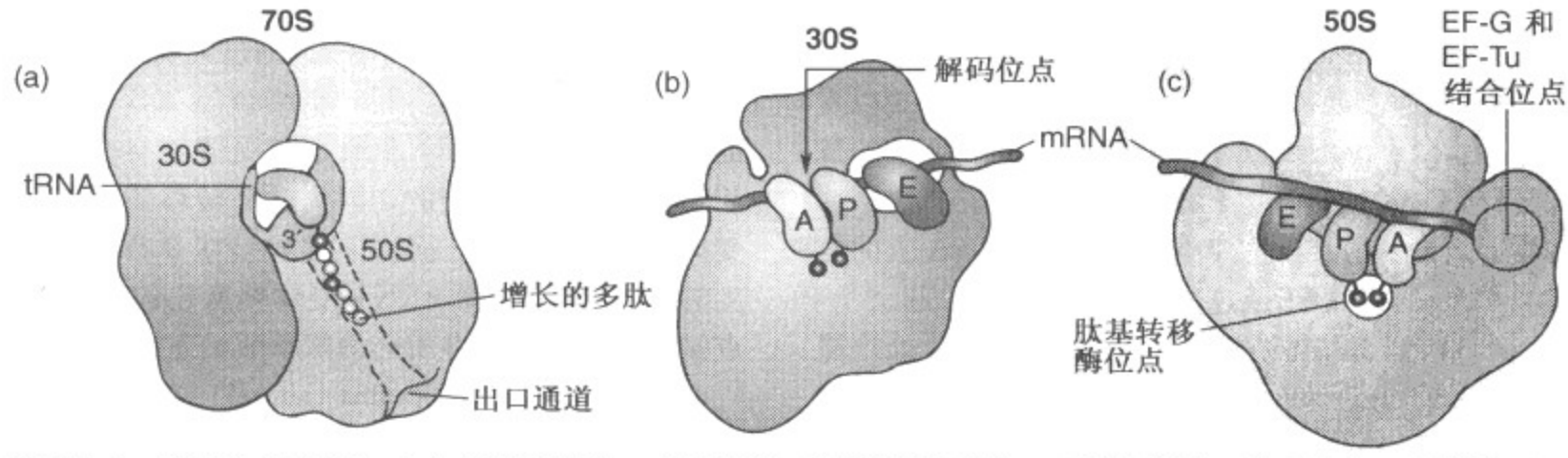


图 19.4　核糖体示意图。（a）70S 核糖体，显示两个亚基间的大洞，一次能容纳 3 个 tRNA。显示位于 P 位的肽酰-tRNA，新生多肽进入 50S 亚基的出口通道。注意，tRNA 通过其反密码子末端与 30S 亚单位相互作用，通过其受体茎与 50S 亚基相互作用。（b）30S 亚基及结合的 mRNA 和 3 个 tRNA。（c）50S 亚基及结合的 mRNA 和 3 个 tRNA。（*Source*：Adapted from Liljas，A.，Function is structure. *Science* 285：2078，1999.）

结构之前，科学家不得不用*E. coli*的生物化学和遗传学数据去解释另一个细菌（嗜热菌）核糖体的结构，在多数情况下这是一种有效方法，但也总有一些怀疑，毕竟两种细菌的生存环境有极大的不同：一个是哺乳动物的肠道，另一个是沸腾的热泉。

最新结构包含了大量数据，尚有待全面分析。然而，已经呈现出几个有趣的发现，最惊人的是晶体的每个单晶都含有两种不同的核糖体结构，命名为“核糖体 I”和“核糖体 II”。两种结构的主要差异是核糖体结构域的刚体运动，其中最显著的是30S 颗粒的头部转动，从“核糖体 I”到“核糖体 II”偏向 E 位点 6°。比较 *T. thermophilus* 和*E. coli* 核糖体 II 时，这种转动更清晰（偏向 E 位点 12°）。

头部的这种转动一定与 mRNA 和 tRNA 在核糖体上的移位有关。实际上，Joachim Frank 和 Rajendra Kumar Agrawal 在 2000 年就做了移位期间核糖体的冰冻-电子显微镜研究，注意到两个亚基的相对运动，在这一过程中，mRNA 的通道变宽以方便移位，移位之后，通道又关闭。因此移位过程中核糖体好像一个棘轮（ratchet），30S 颗粒头部的转动可能是这个棘轮作用的一部分。

真核生物的胞质核糖体比原核生物的复杂得多。在哺乳动物中，整个核糖体的沉降系数是 80S，由 40S 和 60S 亚基组成。真核生物细胞器也有自己的核糖体，但是简单一些，甚至比细菌的核糖体还要简单。

小结 嗜热菌 70S 核糖体结合了 mRNA 类似物和 3 个 tRNA 的复合体晶体结构揭示了以下内容：3 个 rRNA（译者注：应为 tRNA）和多数蛋白质的位置和三级结构都能被确定。位于 A、P、E 位点的 tRNA 的形状和位置是显而易见的。核糖体中 tRNA 结合位点主要是由 rRNA 而不是蛋白质组成的。A、P 位点的 tRNA 反密码子相互间紧密靠近足以和邻近的结合到 30S 亚基上的密码子碱基配对，两个密码子间的 mRNA 扭转了 45°。在 50S 亚基的肽基转移酶口袋内，A、P 位点上 tRNA 的受体茎也相互紧密靠近，仅相距 5Å，这与肽键形成过程中两个受体茎相互作用的需求相一致。可以看到亚基间有 12 个接触点，其中大部分是以 RNA-RNA 相互作用介导的。

*E. coli*核糖体的晶体结构含有两种结构，差异是由核糖体结构域的相对刚体运动引起的。具体来讲，30S 颗粒的头部转动了 6°，相比之下，*T. thermophilus* 核糖体转动了 12°。这种转动可能是移位过程中核糖体做棘轮运动的一部分。

真核生物的胞质核糖体比原核生物核糖体更大、更复杂，但是真核生物细胞器中的核糖体却比原核生物的要小。

核糖体的组成

在第 3 章我们学过*E. coli* 30S 核糖体亚基由 1 分子 16S rRNA 和 21 个核糖体蛋白组成，而 50S 颗粒含有 2 个 rRNA（5S 和 23S）及 34 个核糖体蛋白。用苯酚抽提核糖体去除蛋白质，在溶液中留下 rRNA，因此可相对容易地纯化 rRNA。rRNA 的大小可通过超速离心法确定。

但是核糖体蛋白是非常复杂的混合物，需通过更好的方法来分离。利用单向 SDS-PAGE 可将 30S 核糖体蛋白展现为若干不同的条带，分子质量 8～60kDa，但有些蛋白质在这一方法中不能完全分离。1970 年，E. Kaldschmidt 和 H. G. Wittmann 用双向凝胶电泳几乎完全将两个亚基的蛋白质加以分离。在这个实验中，丙烯酰胺双向凝胶电泳（不含 SDS）仅在两个不同 pH 和丙烯酰胺浓度下进行。

图 19.5 给出了*E. coli* 30S 和 50S 蛋白质的双向电泳结果。每个点含有 1 个蛋白质，对 30S 的蛋白质定名为 S1-S21，对 50S 的蛋白质定名为 L1-L33（L34 看不见）。S 和 L 分别代表小的和大的核糖体亚基。数字编号从最大的蛋白质开始到最小的蛋白质结束，因此 S1 约 60kDa，S21 约 8kDa。从图 19.5 上几乎可以看见所有的蛋白质，并且几乎所有的蛋白质都能与邻近的蛋白质分离开。

真核生物核糖体更加复杂。哺乳动物 40S 亚基含有一个 18S rRNA 和大约 30 种蛋白质。哺乳动物 60S 亚基含有 3 个 rRNA（5S、5.8S 和 28S）以及约 40 种蛋白质。我们在第 10 和 16 章已经学过，5.8S、18S 和 28S rRNA 都来自于同一转录物，由 RNA 聚合酶 I 产生，但 5S rRNA 由 RNA 聚合酶 III 单独转录产生。真核生物细胞器 rRNA 甚至比原核生物的同类 rRNA 还小。例如，哺乳动物线粒体的小核糖体亚基有一个 rRNA，其沉降系数仅为 12S。

小结 *E. coli* 30S 亚基含有 1 个 16S rRNA 和 21 种蛋白质（S1-S21），50S 亚基含有 1 个 5S rRNA，1 个 23S rRNA 和 34 种蛋白质（L1～L34）。真核生物的胞质核糖体比原核生物同类核糖体更大且含有更多的 RNA 和蛋白质。

核糖体的组装

一旦分离出所有的核糖体蛋白，至少在理论上它们就能被加回 rRNA 中产生完全有活性的核糖体。Masayasu Nomura 及同事最先获得了这种重组*E. coli*30S 核糖体亚基。他们将纯化的蛋白质加入 16S rRNA，重新生成了一个有活性的 30S 颗粒，其

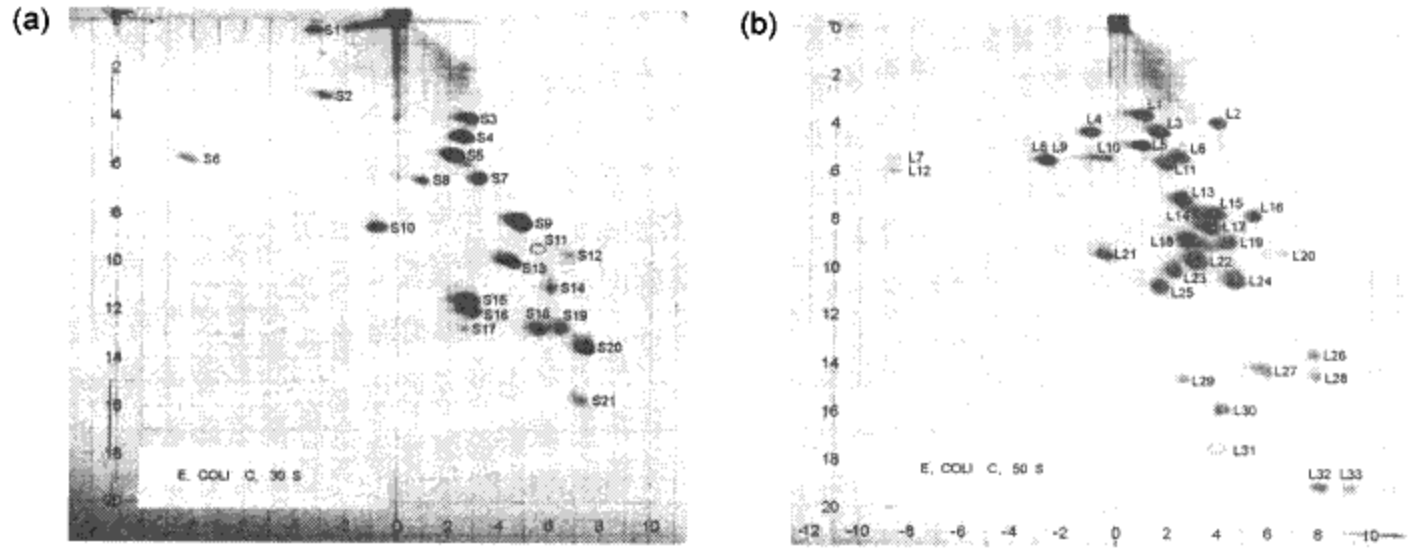


图 19.5　蛋白质的双向凝胶电泳。(a) *E. coli* 30S 亚基 (b) *E. coli* 50S 亚基。蛋白质按编号鉴定，S 表示核糖体小亚基，L 表示核糖体大亚基。第一向（水平方向）电泳：pH8.6，8%丙烯酰胺；第二向（垂直方向）电泳：pH4.6，18%丙烯酰胺。在这两张胶图上看不见蛋白质 S11 和 L31，它们的位置根据其他实验结果用虚线圆点标出。(*Source*: Kaltschmidt, E. and H. G. Wittmann, Ribosomal proteins XII: Number of proteins in small and large ribosomal subunits of *Escherichia coli* as determined by two-dimensional gel electrophoresis. *Proceedings of the National Academy of Sciences USA* 67 (1970) f. 1-2, pp. 1277-78.)

活性检测是将重组的 30S 颗粒与 50S 颗粒结合生成能进行体外翻译的 70S 核糖体。

这种体外重组实验表明 30S 亚基是一种自组装 (self-assembling) 实体，其各组分装配成有活性单位时不需要外来物帮助。这让 Nomura 去探究一个有趣的问题：核糖体各组分以什么顺序组装呢？他通过几种方式回答了这个问题。在其中一个实验中，Nomura 及同事用放射性标记所有的蛋白质，然后各自分别加入 16S rRNA，通过闪烁计数法测定它们的结合（没有 16S rRNA，就无法形成任何颗粒）。S4 和 S8 两种蛋白质与 16S rRNA 结合得非常好（占天然 30S 亚基的 80%～90%），另外 4 种次之，剩下的根本不结合。这表明 S4 和 S8 在所有蛋白质中最先结合。确实，一旦这两个蛋白质结合，下一组蛋白质就跟着结合，之后另一组也可以结合了。

图 19.6 是这类实验的一个例子，Nomura 小组试图确定 S12 结合所必需的蛋白质。图 19.6a 证明 S4、S7、S8、S13、S16 和 S20 这一组蛋白不足以促进 S12 的结合。相反，图 19.6b 显示 S4、S8、S16 和 S17 这一组蛋白质是有效的。图 19.6b 中唯一没有出现在图 19.6a 中的新蛋白质是 S17，表明 S17 在 S12 结合中是重要的，虽然这个实验不能告诉我们 S12

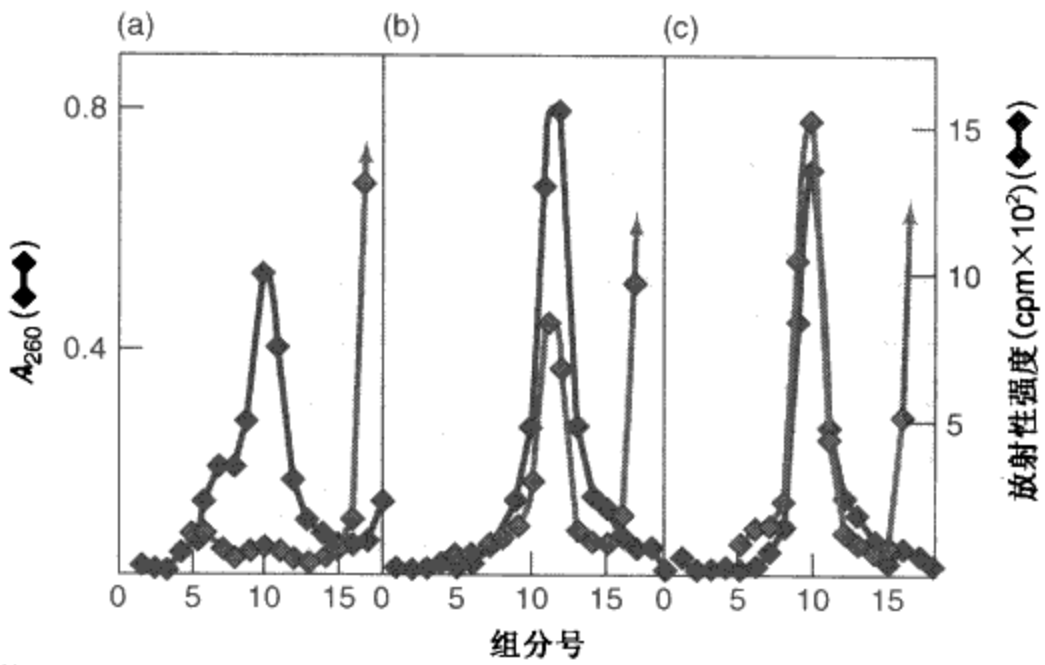


图 19.6　30S 核糖体蛋白对 S12 结合 16S rRNA-蛋白质复合物的影响。Nomura 及同事把 [^{3}H] S12 加到 16S rRNA 中，再加入以下蛋白质混合物：(a) S4、S7、S8、S13、S16 和 S20；(b) S4、S8、S16 和 S17；(c) 所有蛋白，除了 S12。孵育之后，对 3 组蛋白质混合物进行蔗糖密度梯度离心，收集样品，用闪烁计数法测量其放射性并检测 260nm 处的光吸收。观察到对照组 (c) 有最大结合。(b) 组只有 4 种蛋白质，也有相当显著的结合。(a) 组缺少 S17，没有结合。(*Source*: Adapted from Held, W. A., B. Ballou, S. Mizushima, and M. Nomura, Assembly mapping of 30S ribosomal proteins from *Esherichira coli*: Further studies. *Journal of Biological Chemistry* 249: 3109, 1974.)

是否直接与 S17 结合。

按照上述策略，Nomura 建立了如图 19.7 所示的核糖体**组装图**（assembly map）。图中的箭头表示一个蛋白质促进另一个蛋白质的结合。如 S17 促进 S12 的结合。箭头越粗，促进作用越强。如 S17 对 S12 结合的影响比对于 S5 更强。

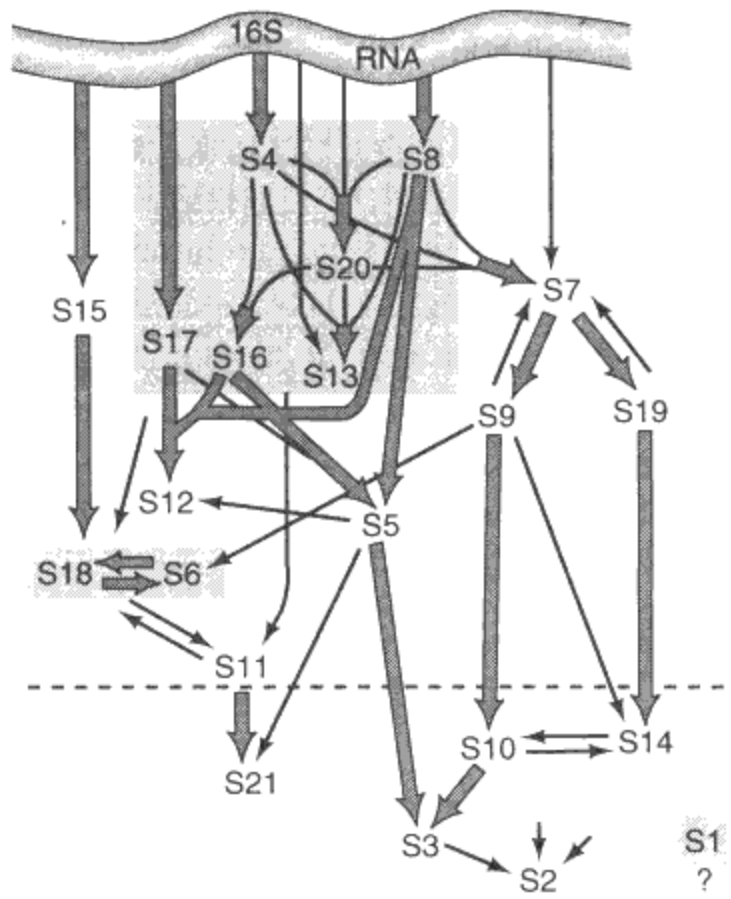


图 19.7　大肠杆菌 30S 核糖体亚基组装图。深色箭头表示强的促进作用，细箭头表示促进作用较弱。从大阴影框出来指向 S11 的箭头表示其中的某些蛋白质能促进 S11 的结合，但还不清楚是哪个。虚线以上的蛋白质在重组中间产物的颗粒（RI）中都有发现，虚线以下的在中间产物中未发现。（*Source*：Adapted from Held, W. A., B. Ballou, S. Mizushima, and M. Nomura, Assembly mapping of 30S ribosomal protein from *Esherichira coli*: Further studies. *Journal of Biological Chemistry* 249: 3109, 1974.）

那么，这个装配图与体内的装配过程有关系吗？证据表明是有关系的。例如，图中最后加入的蛋白质应在核糖体的外面。确实，这些蛋白质对于蛋白水解酶和蛋白质结合试剂的攻击非常敏感。此外，*E. coli*冷敏感突变株在低温下产生只有 9 种蛋白质的缺陷型 30S 颗粒，估计这些蛋白质是体内第一批加入 16S rRNA 的。果然，除了一个例外，其他都是装配图中早期出现的。

50S 亚基的重新装配更具挑战性。首先，*E. coli* 的 50S rRNA 和蛋白质在体外不能自组装，问题出在 50S 亚基的组装过程太依赖温度了，以至于核糖体蛋白在组装所需的高温下会变性。为克服这一难题，实验者转而采用耐高温的嗜热脂肪芽孢杆菌。他们相信这种生活在极高温环境中的细菌应该含有能忍受 50S 亚基重组所需高温的热稳定蛋白。他们是正确的，有活性的嗜热脂肪芽孢杆菌的 50S 颗粒能在体外重新组装，并且这一重组过程的最适温度为 60℃。至 1974 年，用避免极高温的两步实验法成功地在体外重新组装了 *E. coli*核糖体 50S 亚基。尽管有这一早期成功的例子，但目前尚未获得其他任何生物的 50S 核糖体亚基的全面组装图。

小结　30S 核糖体亚基的体外重组从 16S rRNA 开始，蛋白质有顺序地、协作性地结合上去，先加入的蛋白质帮助后面的蛋白质结合于组装颗粒上。

30S 亚基的精细结构

知道*E. coli* rRNA 的序列后，分子生物学家就立刻提出了其二级结构模型，目的是发现在分子内进行最佳碱基配对的最稳定分子。图 19.8 是 16S rRNA 的共有二级结构，该结构已被 30S 核糖体亚基的 X 射线晶体学研究所证实。注意，这个分子存在广泛的碱基配对，并且可被分成 3 个几乎独立折叠的结构域（其中一个具有两个亚结构域），以不同的颜色突出显示。

16S rRNA 的三维排列怎样与完整核糖体亚基中核糖体蛋白的位置相关呢？获得这一信息的最好方法是进行 X 射线晶体学研究。2000 年 V. Ramakrishnan 及同事成功地获得分辨率为 3.0 Å 的嗜热细菌 30S 亚基的晶体结构。几乎在同一时间，Francois Franceschi 小组确定了分辨率为 3.3Å 的同样结构。Ramakrishnan 及同事获得的结构包含 16S rRNA（99%以上的 RNA 分子）的全部有序区域和 20 种核糖体蛋白（占蛋白质总数的 95%）。结构中所缺失的蛋白质部分仅在其无序末端。

图 19.9a 是 16S rRNA 的立体图，RNA 清楚地勾勒出核糖体的所有重要部位，包括头部、平台和基部（这里称为主体）。另外可见连接头部与主体的颈、从头部向左侧伸出的喙（有时称为鼻）以及位于主体左下方的一个刺。颜色标识与图 19.8 相同，以强调 16S rRNA 的二级结构元件与独立的三级结构元件的对应。图 19.9b 显示带有蛋白质的 30S 亚基的正面和背面视图，蛋白质没有引起亚基整体形状的改变。换句话说，蛋白质并没有专门构成亚基的任何主要部分。以上说法并不意味着 16S rRNA 在没有蛋白质情况下就会形成这里所示的形状，只

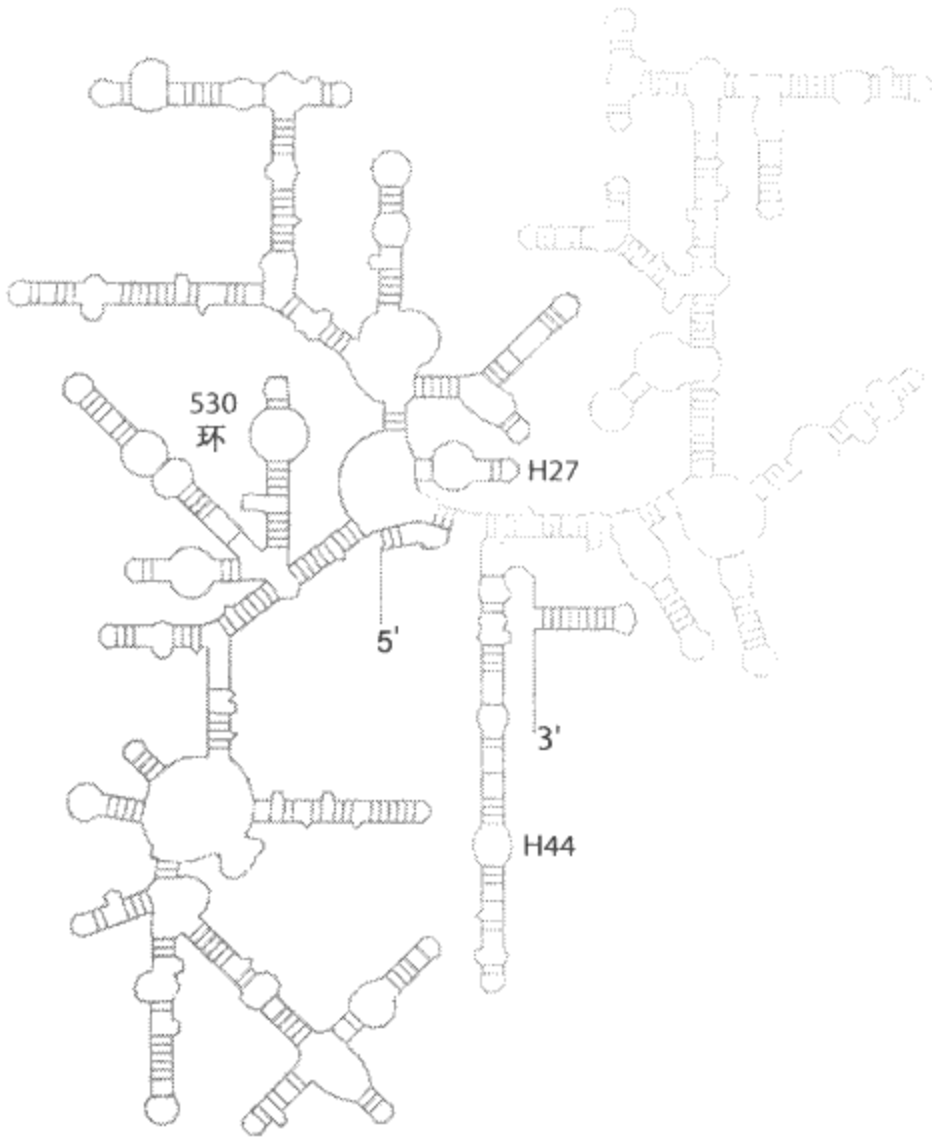


图 19.8　16S rRNA 的二级结构。本结构基于嗜热菌 30S 核糖体亚基的最佳碱基配对和 X 射线晶体学分析。标出的两个螺旋（H27 和 H44）和 530 环在本章的后面讨论。红色，5′结构域；绿色，中心结构域；黄色，3′大结构域；蓝绿色，3′小结构域。（*Source*：Adapted from Wimberly，B. T.，D. E. Brodersen，W. M. Clemons Jr.，R. J. Morgan-Warren，A. P. Carter，C. Vonrhein，T. Hartsch，and V. Ramakrishnan，Structure of the 30S ribosomal subunit. *Nature* 407（21 Sep 2000）f. 2a，p. 329.）

表明 rRNA 是 30S 亚基很重要的部分，以至于在完整亚基中其形状与亚基自身骨架相似。大部分蛋白质的位置与早期用其他方法确定的位置非常一致。

小结　16S rRNA 的序列研究引起对该分子二级结构（分子内碱基配对）的推测。X 射线结晶学研究证实了这些研究的结论。30S 亚基的 16S rRNA 具有广泛的碱基配对，其形态基本上勾勒出整个颗粒的形态。X 射线结晶学研究也证实了大部分 30S 核糖体蛋白的位置。

30S 亚基与抗生素的相互作用　Ramakrishnan 及同事也获得了结合 3 种不同抗生素的 30S 亚基的晶体结构。其中壮观霉素抑制移位，链霉素引起翻译错误，巴龙霉素通过另一种机制增加出错率。基于这些结果连同 30S 亚基自身的结构，我们对翻译机制有了更深入的了解。

首先，他们从 70S 核糖体整体结构上（图 19.1）对 30S 亚基结构上 3 个氨酰-tRNA 进行了定位。图 19.10a 和图 19.10b 显示两种不同视图所示结合在 30S 亚基 A、P、E 位点的氨酰-tRNA 反密码子茎环的位置和假想 mRNA 密码子的位置。令人意外的是，A 位和 P 位的密码子和反密码子位于 30S 亚基颈部附近几乎没有蛋白质的地方。因此，密码子-反密码子识别发生在 16S rRNA 片段和很少量蛋白质围绕的环境中。图 19.10c 显示 16S 的哪些部分参与了这 3 个位点。

了解 30S 亚基上 3 种抗生素的结合位置有助于阐明 30S 亚基的两种活性：移位和解码（decoding，密码子-反密码子相互作用）。30S 亚基的几何学研究提示，移位一定包括头部相对于主体的移动。**壮观霉素**（spectinomycin）是一种阻止移位的刚性三环分子，它在 30S 亚基上的结合位点靠近其移位时头部转动的基点。因此，壮观霉素处于阻断移位所必需的头部转动的位置。

链霉素（streptomycin）　通过干扰密码子-反密码子的初始识别及校正而增加翻译的出错率。链霉素在 30S 亚基上的结合位点（图 19.11）可部分提供其作用机制的线索。链霉素十分靠近 A 位点，而此处发生解码。具体讲，它与 16S rRNA 的螺旋 H27 上的 A913 有紧密接触。

链霉素的这一位置很重要，因为在翻译时 H27 螺旋有两种可变的碱基配对模式，并且这两种模式影响翻译准确性。第一种模式称为 ***ram***（来自 ribosome ambiguity，核糖体含糊）模式。正如其名字所指，这种碱基配对模式能稳定密码子与反密码子甚至非同类反密码子间的相互作用，所以翻译的准确性在 *ram* 状态很低（Ramakrishnan 及同事获得的核糖体晶体结构在 *ram* 状态下具有 H27 螺旋）。另一种碱基配对模式是**严谨的**（restrictive），要求密

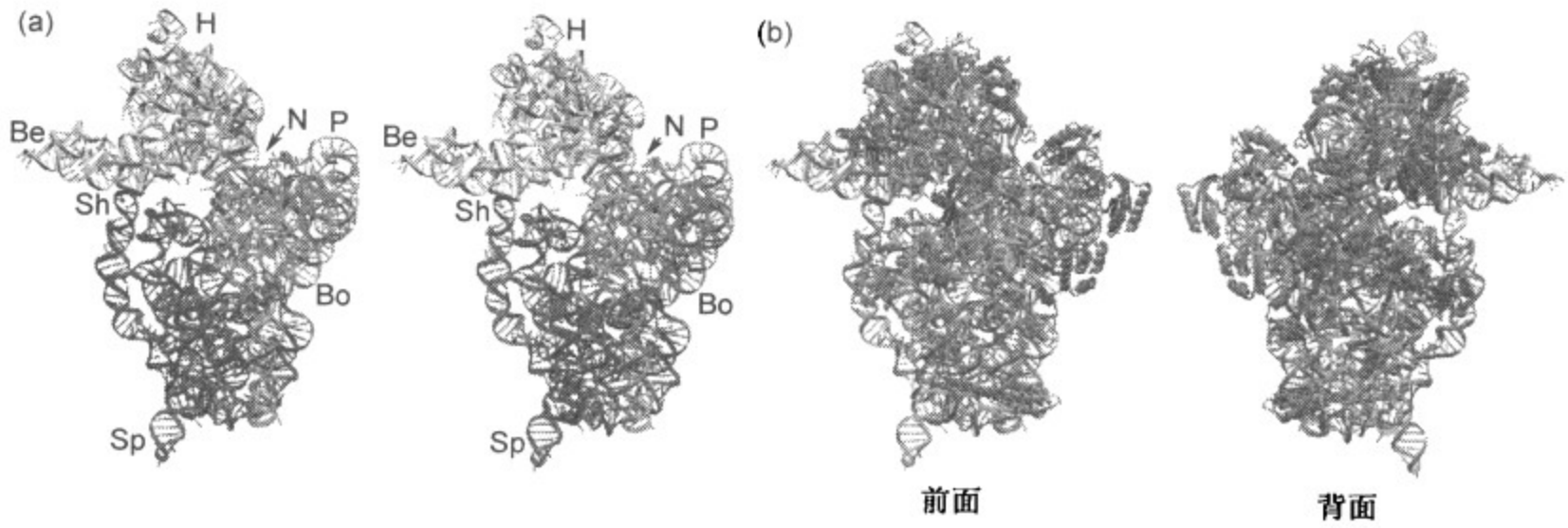


图 19.9　30S 核糖体亚基的晶体结构。(a) 嗜热细菌 30S 亚基的 16S rRNA 部分的立体图，主要特征：H：头部；Be：喙部；Sh：肩部；N：颈部；P：平台部；Bo：主体；Sp：刺。颜色与图 19.8 含义相同。(b) 带有蛋白质的 30S 亚基（紫色）的前面和背面观。前面通常认为是 30S 亚基与 50S 亚基的相互作用面。注意，这是两个不同角度的视图，不是立体图。(*Source*: Wimberly, B. T., D. E. Brodersen, W. M. Clemons Jr., R. J. Morgan-Warren, A. P. Carter, C. Vonrhein, T. Hartsch, and V. Ramakrishnan, Structure of the 30S ribosomal subunit. *Nature* 407 (21 Sep 2000) f. 2b, p. 329. Copyright © Macmillan Magazines Ltd.)

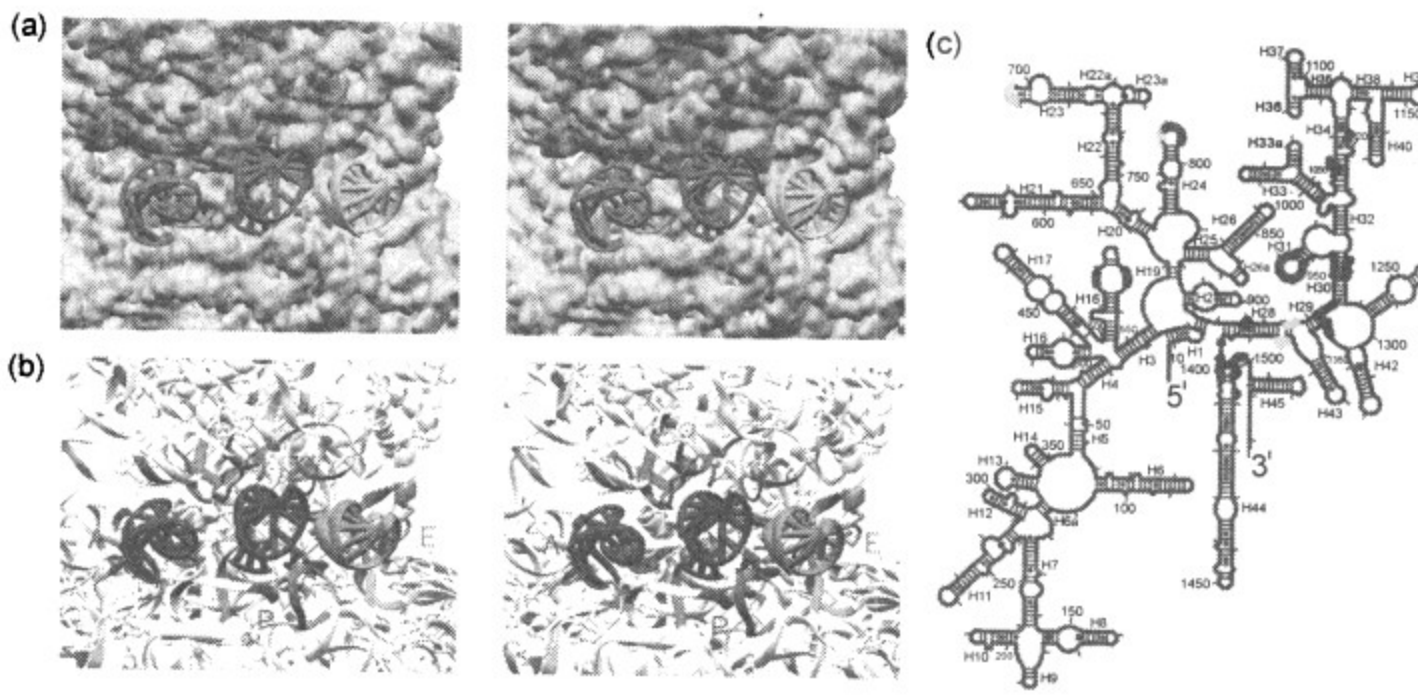


图 19.10　30S 核糖体亚基的 A、P、E 位点的定位。(a) 和 (b) 是 30S 核糖体亚基上推测的反密码子茎环和 mRNA 密码子位置的两个视图。反密码子茎环颜色为紫红色（A 位点）、红色（P 位点）和黄色（E 位点）。mRNA 密码子颜色为绿色（A 位点）、蓝色（P 位点）和虚线紫红色（E 位点）。(c) 16S rRNA 的二级结构，显示各个位点的相关区域，与 (a) 和 (b) 中反密码子颜色一致：紫红色（A 位点）、红色（P 位点）和黄色（E 位点）。(*Source*: Carter, A. P., W. M. Clemons Jr., D. E. Brodersen, R. J. Morgan-Warren, B. T. Wimberly, and V. Ramakrishnan, Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature* 407 (21 Sep 2000) f. 1, p. 341. Copyright © MacMillan Magazines Ltd.)

码子和反密码子间准确配对。如果核糖体被锁在 *ram* 状态，会很容易与非关联氨酰-tRNA 结合，且不能转换成校正所要求的严谨状态，结果导致翻译不准确。如果核糖体被锁定在严谨状态，则是高度准确的，很少出错，但氨酰-tRNA 很难结合 A 位点，所以翻译效率低。

链霉素和 30S 亚基间的相互作用表明它能稳定 *ram* 状态，从两方面降低翻译的准确性。第一，解码过程中链霉素偏爱 *ram* 模式，因而促进密码子与非关联氨酰-tRNA 间的配对；第二，链霉素抑制其向校正所必需的严谨模式的转变。

对核糖体蛋白 S12 的突变既能赋予核糖体链霉素抗性又能赋予链霉素依赖性，而所有这些 S12 突变几乎都位于使 H27 的 908～915 和 H18 的 524～

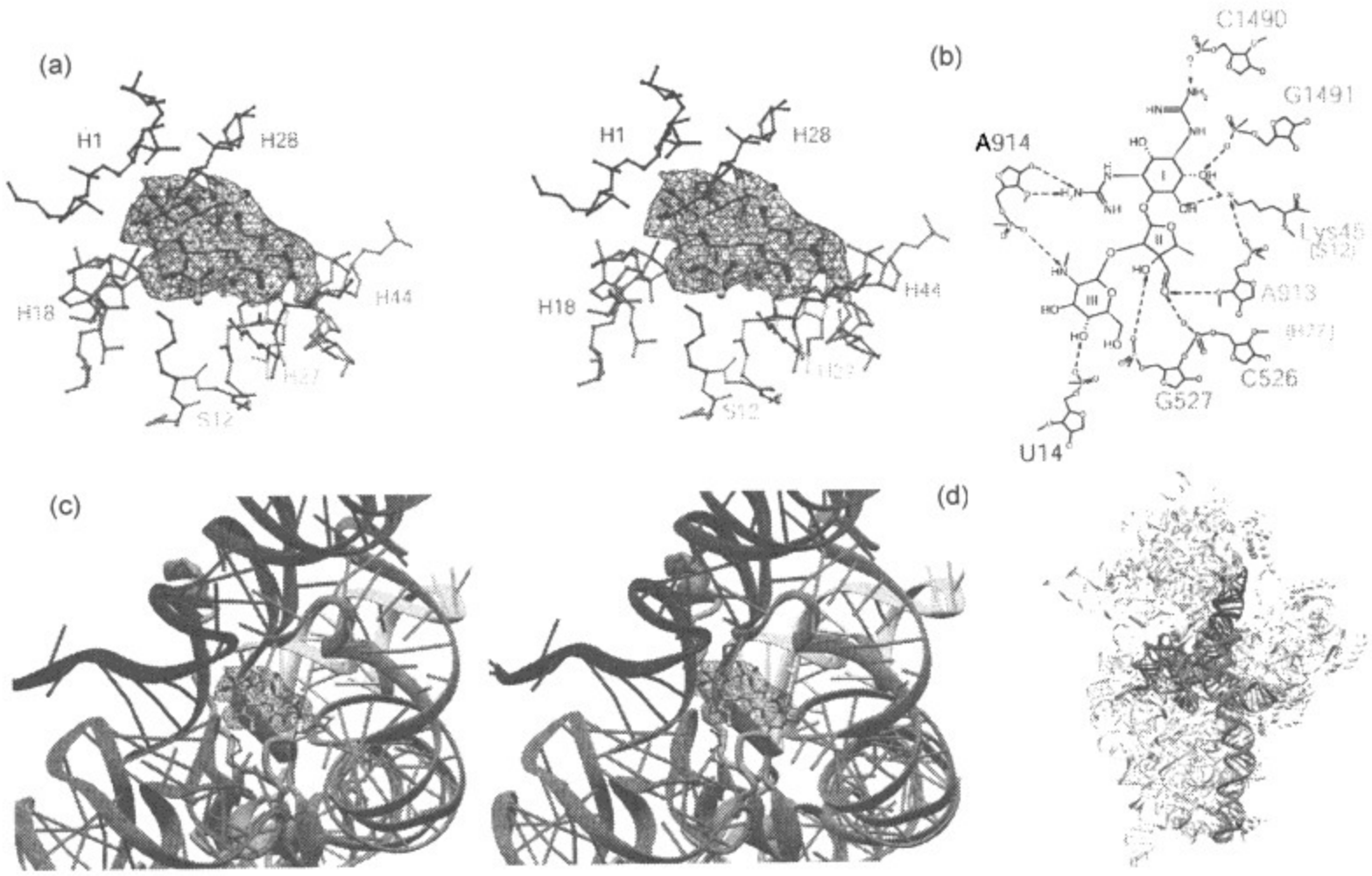


图 19.11　链霉素与核糖体 30S 亚基的相互作用。（a）链霉素与其在 30S 亚基的邻近分子的立体图。链霉素分子显示为被电子密度包围的球-棒模型（实际上，30S 亚基有或无抗生素形成不同的电子密度）。显示 16S rRNA 附近的螺旋，尤其注意 H27 螺旋（黄色），它对于抗生素的活性至关重要，还要注意 A 位点附近的唯一蛋白质 S12（橙色）的位置，它对于链霉素的活性也很重要，红色显示链霉素抗性细胞中被改变的 S12 的氨基酸。（b）链霉素特殊基团（含有标号为 I、II、III 的环）与 30S 亚基上相邻原子的相互作用。注意 H27 的 A913 和 S12 的 Lys45 的相互作用。（c）另一个链霉素与其邻近分子的立体图。颜色含义与（a）图相同。再次注意 H27（黄色）和 S12（橙色）。（d）链霉素结合位点在完整 30S 亚基上的位置。链霉素以小的红色空间充填模型表示，所处位点是各色 16S rRNA 螺旋汇聚之处。（*Source*：Carter，A. P.，W. M. Clemons Jr.，D. E. Brodersen，R. J. Morgan-Warren，B. T. Wimberly，and V. Ramakrishnan，Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature* 407（21 Sep 2000）f. 5，p. 345. Copyright © Macmillan Magazines Ltd.）

527 稳定的蛋白质区域，这里也是 16S rRNA 使 *ram* 状态稳定的地方。

这些现象使 Ramakrishnan 及同事提出了以下的二元假设（two-part hypothesis）：第一，引起链霉素抗性的 S12 突变使 *ram* 状态不稳定，足以抵消由链霉素产生的 *ram* 状态稳定性，结果使核糖体在即使有链霉素存在的情况下也能正常工作。第二，引起链霉素依赖性的 S12 突变对 *ram* 状态的稳定性降低得太多，以至于突变核糖体需要链霉素来维持 *ram* 状态的正常稳定性，结果导致核糖体没有链霉素就不能进行正常的蛋白质翻译。

换句话说，准确有效的蛋白质翻译依赖于核糖体 *ram* 状态和严谨状态间的平衡。链霉素通过促进 *ram* 状态使平衡向低准确性、高效率倾斜，而 S12 突变通过促进严谨状态使平衡向高准确性、低效率倾斜。为验证这一假说，需要有处于严谨状态的 30S 亚基结构，以便与 Ramakrishnan 及同事所提供的 *ram* 状态相比较。

巴龙霉素（paromomycin）　通过结合到 A 位点降低翻译的准确性。Ramakrishnan 及同事在 2000 年发现，巴龙霉素结合于 H44 螺旋的大沟中并使碱基 **A1492** 和 **A1493** “弹出”，即迫使这些碱基从大沟出来，并置于与 A 位点密码子与反密码子间的小沟相互作用的位置。碱基 A1492 和 A1493 是普遍保守的，为翻译活性所绝对必需，其中任一碱基的突变都是致命的。

由此，Ramakrishnan 及同事提出以下假设：在正常解码过程中，碱基 A1492 和 A1493 弹出，并与小沟（由 A 位点的密码子和反密码子碱基配对形成）里糖的 2′-OH 形成氢键，从而有助于稳定密码

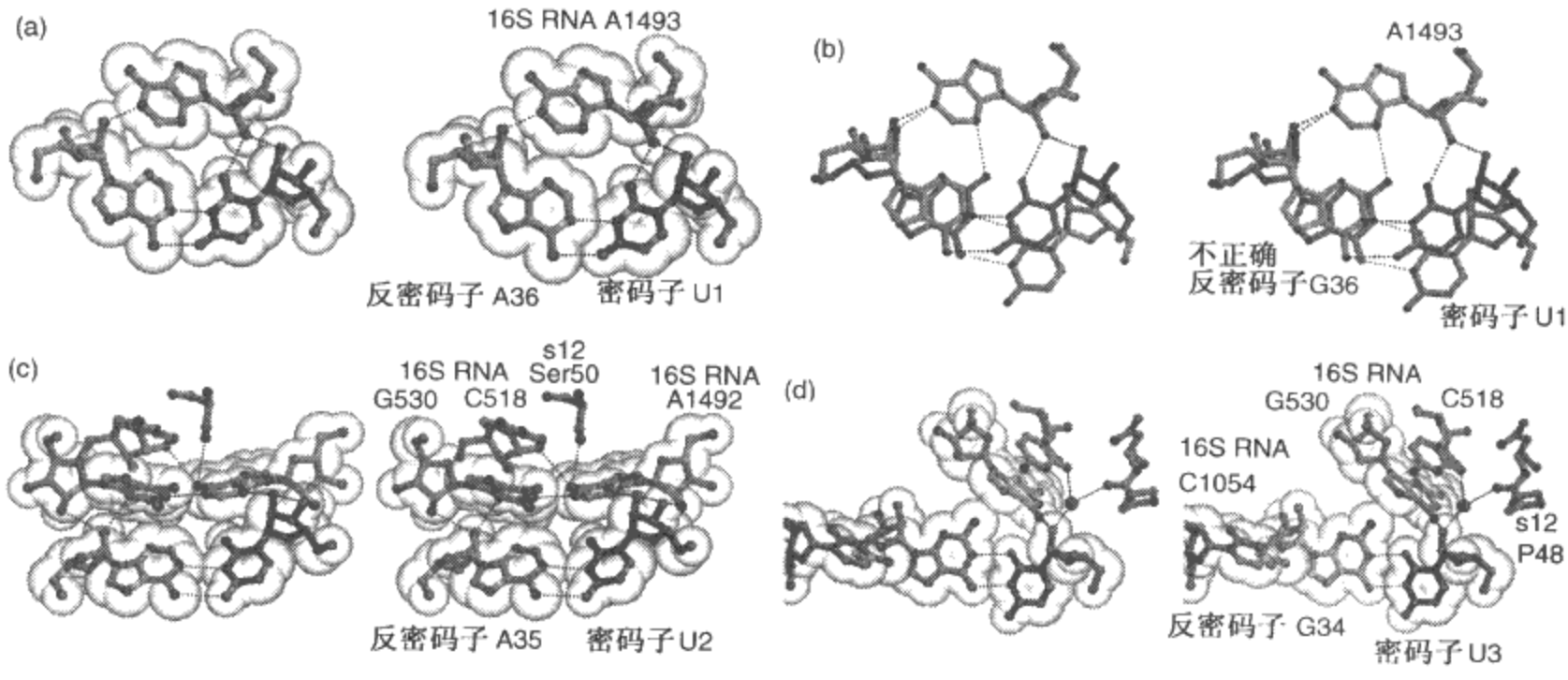


图 19.12　密码子-反密码子碱基对与核糖体 30S 亚基元件间相互作用的立体图。（a）H44 螺旋的 A1493 与 U1-A36 碱基对的小沟结合。（b）与（a）相似，但也显示用 G 取代反密码子中 A36 的结果，所以在 G36 和 U1 间形成了 G-U 摇摆配对。现在 G36（红色）和 U1（蓝紫色）的位置可以与 A36（金色）和 U1（紫色）的正常位置相比较。注意，U1 已经被取代了，因而失去与 A1493（黑色点线）的正常相互作用。这使得相互作用变得不稳定，并有助于核糖体识别同类 A-U 反密码子-密码子碱基对和涉及密码子中第一个碱基的非同类 G-U 反密码子-密码子碱基对。（c）A1492 和 G530 结合到 U2-A35 碱基对的小沟中。（d）摇摆碱基对 U3-G34 通过 U3 与 G530 以及 Mg^{2+}（红紫色球）与 C518 和 S12 蛋白的 48 位脯氨酸而发生相互作用。16S rRNA 的 C1054 底部靠近 G34。（*Source*：From Ogle et al.，*Science* 292：p. 900 © 2001 by the AAAS.）

子与反密码子间的相互作用。这很重要，否则仅靠 3 个碱基对不能保证稳定性。弹出这两个碱基通常需要能量，但是巴龙霉素通过迫使碱基弹出，消除了这种能量需求。因此巴龙霉素稳定了氨酰-tRNA 包括非关联氨酰-tRNA 对 A 位点的结合，增加了出错率。

由于带巴龙霉素的 30S 亚基晶体结构中不存在密码子和反密码子，所以没有直接的证据证明所提出的相互作用，即碱基 A1492 和 A1493 与密码子-反密码子双链小沟的相互作用。

2001 年，Ramakrishnan 及同事提供了证明他们这一假说的直接证据。他们将嗜热菌 30S 亚基的晶体浸在溶液中，该溶液含有对应 $tRNA^{Phe}$ 反密码子茎环的 17nt 的寡聚核苷酸和编码苯丙氨酸的 U_6 寡聚核苷酸，这些分子都很小，足以插入 30S 亚基的适当位置，分别模拟完整氨酰-tRNA 和 mRNA 的反密码子和密码子。

图 19.12 显示这一复合物晶体结构的部分立体视图。图 19.12a 清楚地显示了 H44 螺旋的 A1493 与位于第一个密码子-反密码子碱基对（U1-A36）小沟中的两个核苷酸的糖 2′-羟基接触。图 19.12b 显示如果反密码子的 A36 被 G 替代，就不利于与 A1493 的相互作用。图 19.12c 中，H44 螺旋的 A1492 和 16S rRNA 530 环的 G530 与位于第二个密码子-反密码子碱基对（U2-A35）的两个核苷酸的糖 2′-羟基接触。在密码子-反密码子识别中这是两个最重要的碱基对，两者都通过碱基 A1492 和 A1493 的弹出外加一些其他核糖体元素而稳定。

第三个密码子-反密码子碱基对（摇摆碱基对 U3-G34，图 19.12d）也被核糖体元素所稳定，包括 S12 核糖体蛋白的 G530 和 P48，但不是碱基 A1492 和 A1493。

图 19.13 总结了这些晶体结构所告诉我们的关于 A1492、A1493 的作用和巴龙霉素在密码子-反密码识别中的作用。比较图 19.13a 和图 19.13b 可见，巴龙霉素结合于螺旋 H44 的里面，迫使碱基 A1492 和 A1493 离开螺旋进入 A 位点。图 19.13c 显示无巴龙霉素时密码子-反密码子的识别，也显示了 A1492 和 A1493 占据与巴龙霉素同样的位置，这两个 rRNA 的碱基处在理想的位置来感知第一个和第二个碱基对中碱基间的配合，这种感知是通过感觉位于密码子-反密码子双螺旋小沟的核糖的位置实现的。图 19.13d 显示与巴龙霉素存在时相同的结构，再次显示没有抗生素时结构几乎没有变化。

所有这些发现都与以下假设一致，即巴龙霉素通过将 A1492 和 A1493 推出螺旋 H44，补偿了核糖

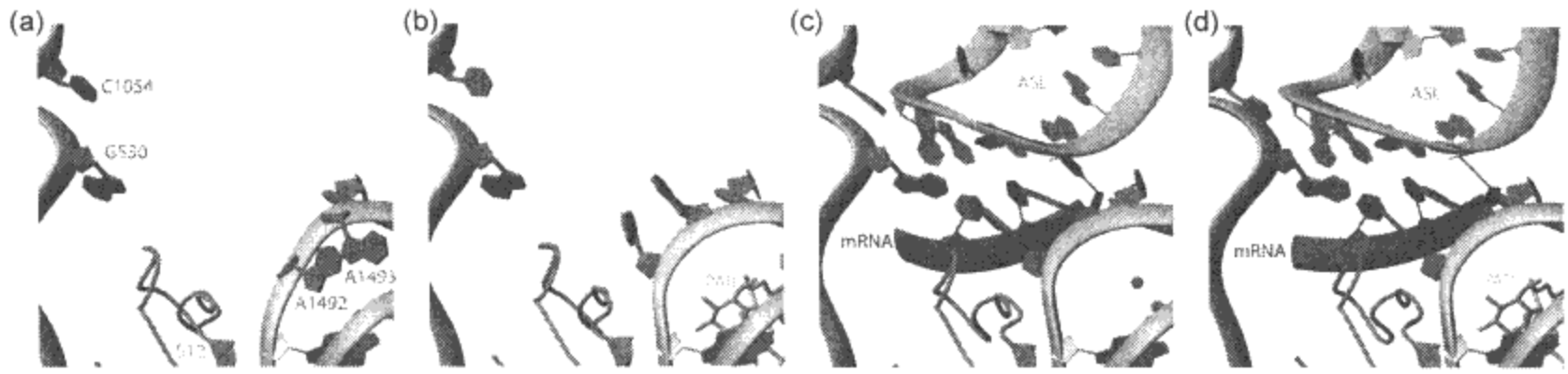


图 19.13　A 位点在存在或缺少 tRNA、mRNA 及巴龙霉素时的局部结构。(a) A 位点的本身结构，注意在 H44 螺旋中 A1492 和 A1493 的位置，这些碱基的位置很易变；(b) 存在巴龙霉素时 A 位点的结构，巴龙霉素结合到 H44 螺旋内部迫使 A1492 和 A1493 处于螺旋外面的位置并进入 A 位点；(c) A 位点的结构，有 mRNA 和 A 位点 tRNA 反密码子茎环 (ASL)，推测碱基 A1492 和 A1493 在 A 位点的位置同巴龙霉素单独存在时一样；(d) 与图 (c) 相同，但有巴龙霉素。抗生素没有引起差异，因为 A1492 和 A1493 已经和 A 位点作用了。(*Source*: From Ogle et al., *Science* 292: p. 900. © 2001 by the AAAS.)

体上诱导密码子和反密码子间配合的部分能量消耗。这样一来，抗生素使非关联密码子和反密码子间的配对更容易，因此增加了 mRNA 误读的频率。

小结　30S 核糖体亚基起两个作用，它促进密码子和氨酰-tRNA 反密码子间正确解码和校正，还参与蛋白质翻译。利用 30S 核糖体亚基带有 3 种抗生素的晶体结构可使翻译和解码过程清楚地显现出来，这些抗生素干扰 30S 亚基的这两种作用。壮观霉素结合于 30S 亚基的颈部附近，在那里干扰移位所必需的头部移动。链霉素结合于 30S 亚基的 A 位点附近，并稳定核糖体的 *ram* 构型，使非正确氨酰-tRNA 相对容易地结合于 A 位点，并通过阻止向校正所必需的严谨构型的转变而降低翻译的忠实性。巴龙霉素结合于 16S rRNA 的 H44 螺旋 A 位点附近的小沟内，使碱基 A1492 和 A1493 弹出，从而稳定密码子和反密码子之间的碱基配对。这一弹出过程通常需要能量，但巴龙霉素可迫使这一过程发生，并且使稳定的碱基保持在其位置上。A 位点的这种状态稳定了密码子-反密码子之间的相互作用，包括非关联密码子和反密码子间的相互作用，因此翻译的忠实性降低。

30S 亚基与起始因子的相互作用　我们已经在第 17 章学过 IF1 因子帮助其他起始因子行使功能。IF1 因子的另一个可能作用是防止氨酰-tRNA 结合于 A 位点，直至起始阶段结束。这种对 A 位点的封锁推测具有两种功能。第一，直到 50S 颗粒加入起始复合物后，EF-Tu 指导的 A 位点上的氨酰-tRNA 校正才能发生。因此，对 A 位点的封锁可阻止氨酰-tRNA 的不正确结合，提高翻译的忠实性。第二，它保证起始氨酰-tRNA 结合于 P 位点而不是 A 位点。

Ramakrishnan 及其合作者已经确定了结合于嗜热菌 30S 核糖体亚基上的 IF1 因子的晶体结构（图 19.14）。图 19.14b 和 19.14c 清楚地显示了 IF1 因子结合并封闭了 30S 亚基的 A 位点，占据 tRNA 对 A 位点结合的大部分区域。

他们所研究的晶体结构不包括 IF2，但从第 17 章可以知道，IF1 辅助 IF2 将 fMet-tRNA 结合到 P 位点，而且还知道 IF1 和 IF2 可相互作用。因此很可能 IF1 对 A 位点的结合可帮助 IF2 结合 30S 亚基，就像促进 fMet-tRNA 对 P 位点的结合一样。

最后，由第 17 章知道，IF1 促进两个核糖体亚基的解离。实际上，它也帮助两个核糖体亚基再结合，所以没有改变两者间的平衡。只有在阻止再结合的 IF3 的帮助下，IF1 才表现为核糖体解离的中介。图 19.14 的结构显示 30S 亚基中 16S rRNA 的螺旋 H44 和 IF1 之间的密切接触。已知螺旋 H44 与 50S 核糖体亚基有广泛接触，Ramakrishnan 及其合作者推测 IF1 与螺旋 H44 之间的接触干扰了螺旋 H44 的结构，使其处于核糖体亚基的结合与解离间的过渡态，由此可解释 IF1 如何促进核糖体的结合与分离。

小结　结合于 30S 核糖体亚基的 IF1 的 X 射线晶体结构表明 IF1 因子结合于 A 位点，明显阻断了 fMet-tRNA 与 A 位点的结合，也可能通过推测的 IF1 和 IF2 之间的相互作用积极地促进 fMet-tRNA 结合于 P 位点。IF1 因子也和 30S 亚基的螺旋 H44 密切作用，由此可解释 IF1 因子如何同时促进核糖体亚基的结合与分离。

50S 亚基的精细结构

2000 年，Peter Moore 和 Thomas Steitz 及同事在核糖体结构研究和 X 射线衍射技术领域达到了一个新的里程碑，他们在 2.4 Å 分辨率上确定了 50S 核糖体亚基的晶体结构。研究是在原始古菌 *Haloar-*

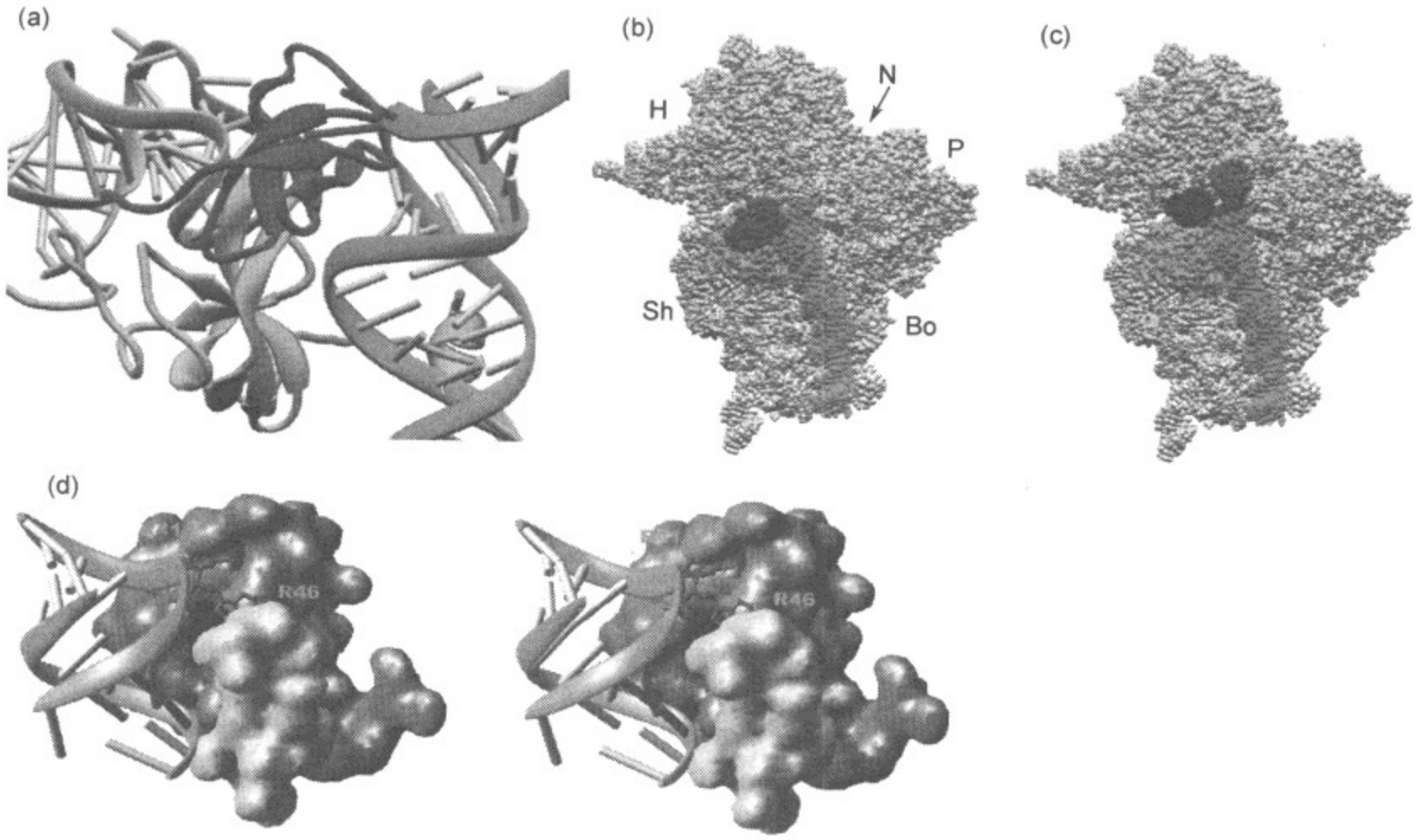


图 19.14　IF1-30S 核糖体亚基复合物的晶体结构。(a) 放大图，IF1 是红紫色，16S rRNA 的螺旋 H44 是青绿色，16S rRNA 530 环是绿色，S12 是橙色。(b) IF1-30S 核糖体亚基复合物的全貌。颜色表示与 (a) 图一样，30S 亚基的其余部分是灰色。(c) 减去 IF1 的全貌，显示 tRNA 在 A 位点（紫色）、P 位点（橘红色）、E 位点（黄绿色）的位置，其他颜色同 (a) 图。注意 A 位点 tRNA 与 (a) 图中 IF1 位点的重叠。(d) 螺旋 H44 的 A1492 和 A1493 碱基与 IF1 和 S12 形成的凹槽间的紧密结合。R41 和 R46（蓝色）是 IF1 与 A1492 和 A1493 相作用的精氨酸，其他颜色同 (a) 图。(*Source*: From Carter et al., *Science* 291: p. 500. © 2001 by the AAAs.)

cula marismortui 的 50S 亚基上进行的，因为能够从这种微生物中制备适合 X 射线衍射的 50S 亚基晶体。结构见图 19.15，包括亚基中 rRNA 的 3045 个核苷酸中的 2833 个（5S rRNA 的全部 122 个核苷酸）和 31 个核糖体蛋白中的 27 个。其他蛋白质排列不规则，不能准确定位。

两个亚基的明显差别在于它们 rRNA 的三级结构，推测 30S 亚基的 16S rRNA 为 1 个有 3 个结构域（three-domain）的结构，而 50S 亚基的 23S rRNA 是结构域间无明显界线的整块结构。Moore 及同事推测造成这一差别的原因是 30S 亚基的结构域相互之间需要相对移动，而 50S 亚基的结构域不需要这样的相对移动。

图 19.15 中的小结构显示了 50S 亚基中蛋白质的位置。如我们在本章前面看到的那样，通常在两个亚基的接触面上缺乏 50S 亚基的蛋白质，特别是肽基转移酶活性位点的中心部位。这是一个令人兴奋的发现，因为关于肽基转移酶活性位点是在 50S 亚基的 RNA 上还是蛋白质上的问题还有一些不确定性（第 18 章）。

要确定是否有蛋白质存在于肽基转移酶活性位点，必须先在晶体结构中找到活性位点。为此，Moore 及同事将 50S 亚基的晶体与两种不同肽基转移酶底物的类似物混合溶解（soak），然后进行 X 射线衍射并计算电子差异图谱，将电子密度与底物类似物相对应，从而确定活性位点。有一个类似物（CCdAp-嘌呤霉素）由 Michael Yarus 设计，模仿肽基转移酶反应过程中的过渡态或中间态，称为“Yarus 类似物”。

图 19.16 显示，Yarus 类似物处于 50S 亚基表面的裂缝内，正好在预计的活性位点，周围没有蛋白质只有 RNA。在另一个类似物上也观察到同样的情况。图 19.17 是一个移去所有 RNA 的活性位点模型，我们可以看到蛋白质到 Yarus 类似物的距离，Yarus 类似物处于酶活中心过渡状态的正中间，对应于正四面体碳原子。最近的蛋白质是 L3，它距离活性中心超过 18 Å，这个距离太远以至于它不能在催化反应中起直接作用。

如果活性中心缺乏蛋白质，那么 RNA 一定具有酶活性。该晶体结构揭示了对应于大肠杆菌 A2451

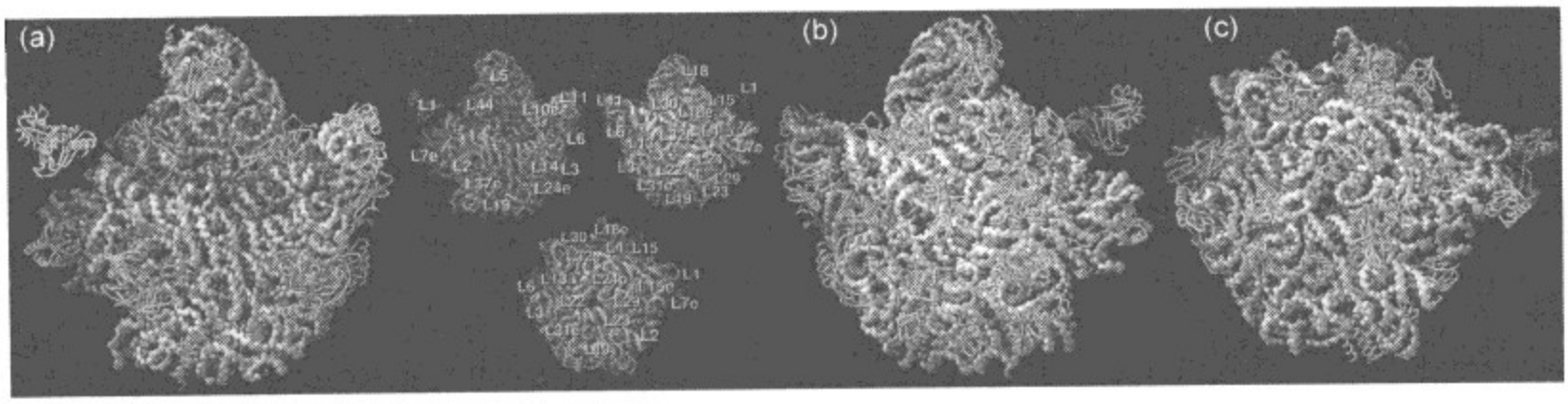


图 19.15　*Haloarcula marismortui* 50S 核糖体亚基的晶体结构。3 个大结构显示亚基的 3 个不同方向。(a) 正面观，或“王冠”图（得此名是因为很像一个三尖王冠）；(b) 背面观（王冠图旋转 180 度）；(c) 底面观，显示在中心的多肽出口通道的末端。RNA 灰色，蛋白质金黄色。左下角的 3 个小结构图和大图的方位一样，标明了蛋白质。有些数字后面的字母“e”表示该蛋白质只在真核生物（而不是细菌）中有同源蛋白。(*Source*: Ban, N., P. Nissen, J. Hansen, P. B. Moore, and T. A. Steitz, The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* 289 (11 Aug 2000) f. 7, p. 917. Copyright © AAAS.)

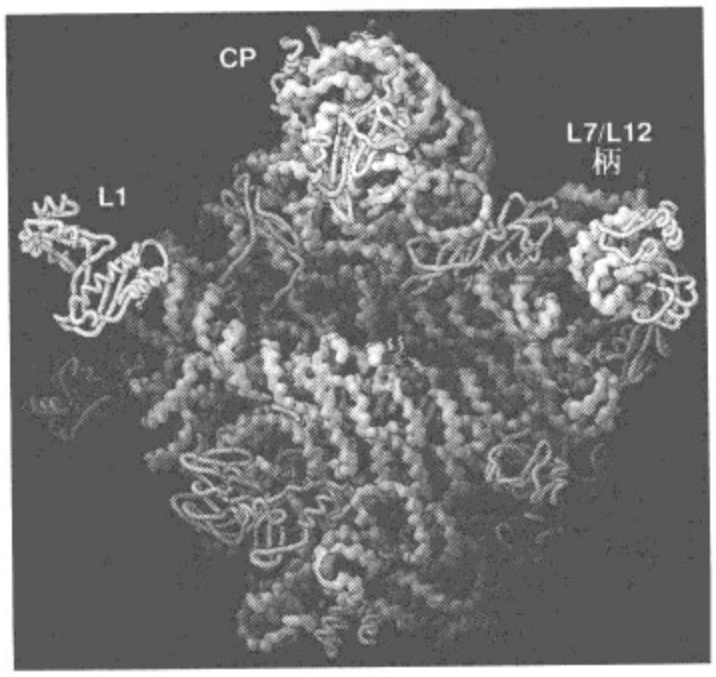


图 19.16　肽基转移酶活性位点的定位。此为 50S 亚基的正面观，同图 19.14，Yarus 类似物的位置应处于肽基转移酶的活性位点（绿色）。注意活性位点周围没有蛋白质（金黄色）。(*Source*: Ban, N., P. Nissen, J. Hansen, P. B. Moore, and T. A. Steitz, The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* 289 (11 Aug 2000) f. 2, p. 907. Copyright © AAAS.)

的腺嘌呤 2486（A2486），最接近位于活性中心的四面体碳原子。这个碱基在检测过的所有三界生物的核糖体中都是保守的，说明该碱基起关键作用。此外，抑制肽基转移酶的氯霉素和碳霉素可结合或接近大肠杆菌的 A2451，而 A2451 突变的大肠杆菌具有氯霉素抗性，这进一步表明该碱基参与了肽基转移反应。

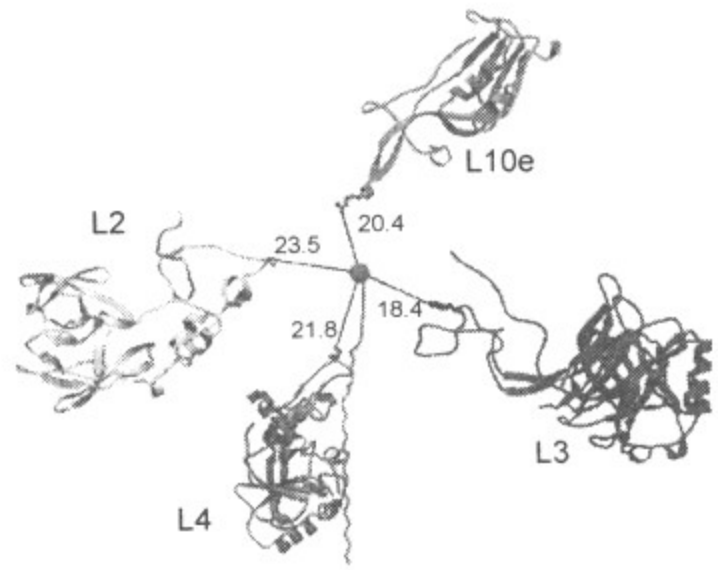


图 19.17　移去所有 RNA 的肽基转移酶活性位点。Yarus 类似物在酶活性中心（暗红色）磷酸化，并携带一根正在延伸的长肽链尾。图中给出了 4 个离活性中心最近的蛋白质，并用最精确的测量方法（Å）分别测出了它们与活性中心的距离。(*Source*: Nissen, P., J. Hansen, N. Ban, P. B. Moore, and T. A. Steitz, The structural basis of ribosome activity in peptide bond synthesis. *Science* 289 (11 Aug 2000) f. 6b, p. 924. Copyright © AAAs.)

假如这个模型是正确的，那么可以预见在 A2486 上的突变会使肽基转移酶活性降低几个数量级。Alexander Mankin 及同事在 2001 年检验了这一推测。他们用分离的蛋白质和 23S rRNA 重组了一个嗜热水生菌的 50S 亚基，这个 23S rRNA 的 A2451（相当于 *H. marismortui* 菌的 A2486）有全部 3 种可能的突变。然后用包括第 18 章介绍的片段反应等 4 种不同分析方法检测了重组 50S 亚基的肽

基转移酶活性。结果显示，没有一个突变引起活性的剧烈降低，在至少一次实验中，23S rRNA 的各种突变至少维持 44%的野生型活性。

如果不是 A2486 在肽基转移酶反应中起主要催化作用，那么是什么在起作用？Scott Strobel 及同事在 2004 年提供的证据表明，P 位点肽酰-tRNA 末端腺苷的 2′-羟基可能在起作用。图 19.18 显示了这个 2′-OH 和 A 位点的氨基酸，羟基对连接肽到 P 位点的碳酰基的碳原子进行亲核性攻击，结果使 P 位点的肽连接到 A 位点的氨酰-tRNA 上，这叫做**转肽反应**（transpeptidation），由肽基转移酶催化。很明显，该 2′-OH 处于一个很好的位置可以通过与氨基上的一个质子形成氢键而发挥作用，使氨基的氮原子成为一个比较好的亲核剂。

如果这个假设正确，除去肽酰-tRNA 的末端腺苷（A76）2′-位的氧原子应该会削弱肽基转移酶活性。Strobel 及同事用两种方法验证了这个推测，即用氢原子（2′-脱氧腺苷）或氟原子（2′-脱氧，2′-氟腺苷，fA）取代该 2′-羟基。当他们对 P 位点 tRNA 的末端腺苷分别做这两种取代时，肽基转移酶活性都被极大地抑制了。

他们的分析是这样的，将［^{35}S］fMet-tRNA 加到 P 位点，然后把 Lys-tRNA 加到 A 位点。在实验中分别以 3 种末端腺苷形式加入 Lys-tRNA：正常的、dA、fA。然后加入肽基转移酶发生一轮移位，使［^{35}S］fMet- Lys-tRNA 位于 P 位点。这为加入嘌呤霉素和观察标记的肽酰-嘌呤霉素从核糖体释放的速度提供了平台。由于嘌呤霉素可以十分迅速地结合到 A 位点，在肽酰-嘌呤霉素释放过程中肽基转移

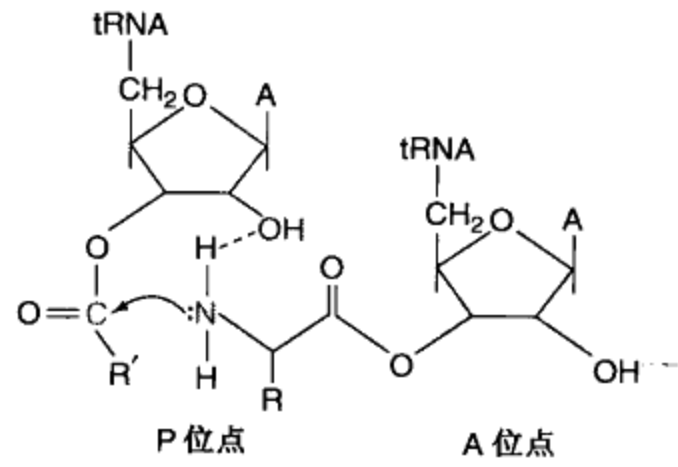


图 19.18　肽基转移酶反应期间在 A 位和 P 位的 tRNA 的位置。P 位点 tRNA 的 2′-OH 为红色，A 位点氨酰-tRNA 的 N 原子为绿色，P 位点肽酰-tRNA 的碳酰基的碳原子为蓝色。注意 P 位点 tRNA 的 2′-OH 对攻击性 N 原子的接近程度。

酶是限速性的，因此释放量可作为肽基转移酶速率的测定。Strobel 及同事利用薄层电泳将释放的肽酰-嘌呤霉素与其他标记的物质分离开，通过磷屏成像测定了放射性。

图 19.19 是实验结果，以正常 RNA 为底物时，肽基转移酶反应在第一个时间点就完成了（10s）。但是以修饰的 RNA 为底物时，反应基本上不发生，即使在 24h 之后也无反应。因此，用氢原子或氟原子取代 P 位点 tRNA 的 2′-羟基都完全阻断了肽基转移酶反应，有力地证明该 2′-羟基是反应所必需的。以这 3 种底物与 A 位点正常 Phe-tRNA 而不是嘌呤霉素反应时也观察到同样的结果，进一步证明了 2′-羟基的重要性。

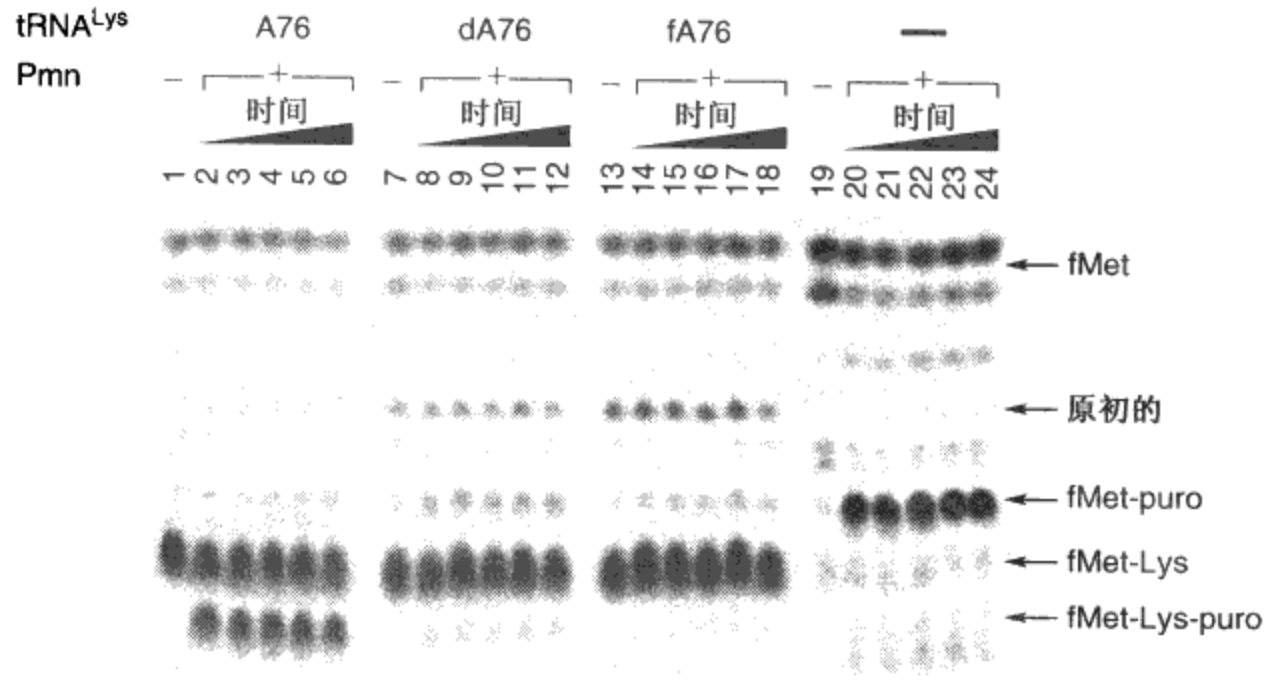


图 19.19　带有经修饰的 tRNA 的肽基转移酶的活性。Strobel 及同事利用 P 位点标记的二肽酰-tRNA 和加到 A 位点的嘌呤霉素进行肽基转移酶反应，P 位点的 tRNA 分别含有正常的 A76、dA76、fA76、无修饰 fMet-tRNA（—），如图上所示。在有或无嘌呤霉素条件下记录反应不同时间的结果（10s、1min、6min、1h、24h）。用薄层电泳分离标记的二肽酰-嘌呤霉素（fMet-Lys-puro），电泳图进行磷屏成像。只有带正常 A76 的 P 位点 tRNA 能够测定肽基转移酶活性。（*Source*：Reprinted from *Nature Structural & Molecular Biology*，vol 11，Joshua S. Weinger，K. Mark Parnell，Silke Dorner，Rachel Green & Scott A. Strobel，"Substrate-assisted catalysis of peptide bond formation by the ribosome，" Fig. 3a，p. 1103，Copyright 2004，reprinted by permission from Macmillan Publishers Ltd.）

但是这些研究仍然未回答 23S rRNA 中高度保守的 A2451（用的是在*E. coli*中的编号）的作用。为探讨这个问题，Norbert Polacek 及同事设计了一种方法，不仅改变碱基的性质，也改变 A2451 糖的性质。当他们从 A2451 中除去腺苷碱基时产生一个非碱基位点（abasic site），利用嘌呤霉素释放分析，发现肽基转移酶活性几乎没有变化。但是，除去 A2451 的 2′-羟基时，肽基转移酶活性几乎降低到 1/10。同时除去 A2451 和 2′-羟基时，肽基转移酶活性完全丧失。相反，在邻近的核苷 A2450 做同样改变时，对活性只有较弱影响，再次强调了 A2451 的特殊重要性。

23S rRNA 的 2451 位 2′-OH 缺失引起的核糖体活性丢失可能是因为对 P 位点 tRNA 较低的亲和性，如果是这样，那么提高 fMET-tRNA 的浓度应该增强活性，但是并没有。那么这个羟基的作用到底是什么呢？刚才研究过的关于 P 位点 tRNA 的 2′-羟基参与转肽反应的证据是充分的，但是 A2451 的 2′-羟基也有可能以这种方式参与反应，其羟基也许通过将反应物正确地定位在活性位点而参与催化反应。

增长的多肽可能通过 50S 亚基的一个通道从核糖体出来。Moore、Steitz 及同事的研究已使这方面相当清楚了。图 19.20 是一个显示出口通道的劈开的 50S 亚基，标记出肽基转移酶中心，通道中示意性地有一个多肽。通道的平均直径为 15 Å，两处窄的地方为 10 Å，刚好能容下一个 α 螺旋蛋白，所以新生多肽的任何进一步折叠都是不可能的。通道壁大部分由亲水性 RNA 构成，新生多肽暴露的疏水基团不会因在通道壁中发现可结合的对象而阻碍脱出过程。

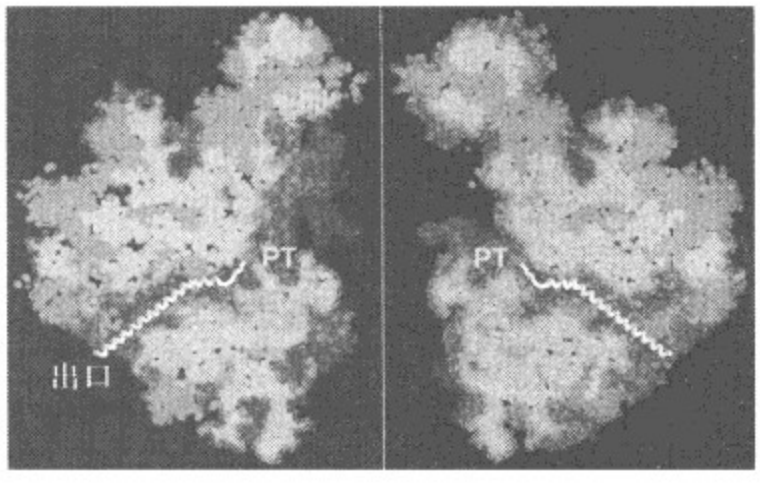


图 19.20　多肽脱出通道。50S 亚基画的像一个从中间劈开的水果。显示出从肽基转移酶（PT）位点开始延伸的出口通道。通道中白色 α 螺旋表示一个正在离开的新生多肽。（*Source*：Ban, N., P. Nissen, J. Hansen, P. B. Moore, and T. A. Steitz, The structural basis of ribosome activity in peptide bond synthesis. *Science* 289 (11 Aug 2000) f. 11a, p. 927, Copyright © AAAs.）

小结　50S 核糖体亚基的晶体结构已在 2.4 Å分辨率上确定。这一结构揭示在核糖体亚基的接触面上相对来说较少有蛋白质，用过渡态类似物在肽基转移酶活性中心 18 Å范围内没有标记到蛋白质。P 位点 tRNA 的 2′-OH 处于最佳位置，与 A 位点的氨酰-tRNA 形成一个氢键，从而协助催化肽基转移反应的进行。与这一假设一致的是除去该羟基几乎消除了肽基转移酶的全部活性。类似地，除去 23S rRNA 的 A2451 的 2′-OH，可强烈抑制肽基转移酶活性，因此该羟基可能也通过形成氢键参与催化反应，或通过帮助反应物正确定位参与催化反应。穿过 50S 亚基的出口通道仅够通过一个 α 螺旋蛋白，通道壁由 RNA 构成，其亲水性允许暴露出疏水侧链的新生蛋白质容易地滑过。

多聚核糖体

在前面的几章我们已经看到一次可以有一个以上 RNA 聚合酶转录一个基因。对于核糖体和 mRNA情况也是一样。实际上，在任何给定时间许多核糖体成串穿过同一个 mRNA 的现象是很普遍的，结果形成一个**多聚核糖体**（polyribosome 或 polysome），如图 19.21 所示。该多聚核糖体中，可以数出 74 个核糖体同时翻译一个 mRNA。通过察看新生蛋白质链可分辨出多聚核糖体的两端，当核糖体从 5′端（翻译开始）向 3′端（翻译终止）行进时，蛋白质链延伸变长。因此，5′端在左下方，3′端在右下方。

在形成真核生物多聚核糖体的过程中，加载到 mRNA 上的第一个核糖体在“开拓”翻译时面临最困难的任务。来自细胞核的 mRNA 带着一些蛋白质，有些是剪接和多聚腺苷酸化过程剩下的，有些是 mRNA 结合蛋白，指导和护送 mRNA 从细胞核出来不被降解。但在两个亚基之间连 mRNA 自己几乎都没有地方，所以当 mRNA 像穿线一样挤进第一个核糖体时，这些蛋白质必须被剥离，并立刻被翻译过程所需的其他蛋白质所替代。

图 19.21 中的多核糖体来自真核生物（一种蚊虫）。因为真核细胞转录和翻译在不同部位发生，所以多聚核糖体总是独立于基因出现在细胞质里。原核生物也有多聚核糖体，但是由于给定基因的翻译和转录在同一时间和同一地点发生，所得图像很复杂，可以看到新生 mRNA 在同一时间合成并被核糖体翻译。图 19.22 显示了大肠杆菌中的这样一种情形。可以看见两个细菌染色体从左到右平行展开，只

有上面的那个片段被转录。从图中可以分辨出转录是从左向右进行的，因为沿这个方向移动时多聚核糖体不断变长，有更多的空间容纳更多的核糖体。注意不要被图 19.21 和 19.22 比例上的差异而误导，在图 19.22 中核糖体看起来较小且新生的蛋白质链看不见，穿过图中的那条链是 DNA，而在图 19.21 中的是 mRNA。在图 19.22 中 mRNA 或多或少是垂直的。

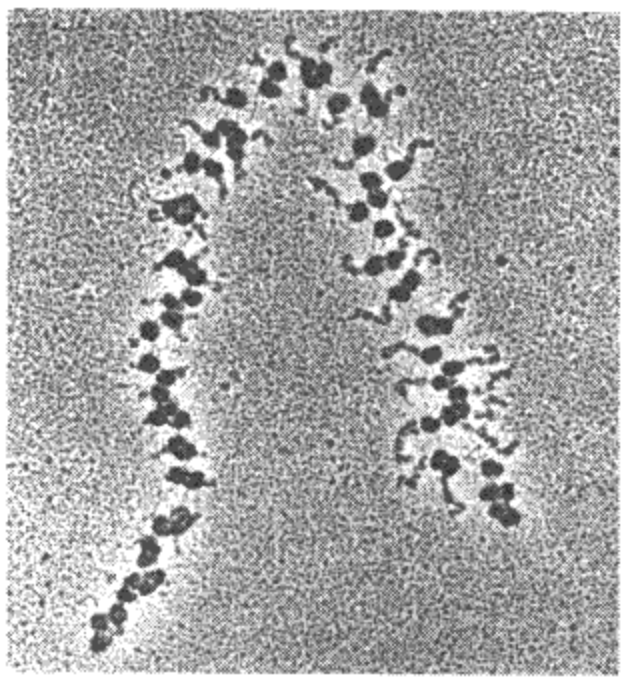


图 19.21　摇蚊 *Chironomus* 多聚核糖体的电镜照片。mRNA 的 5′端在左下角，mRNA 向上弯曲后再向下，3′端在右下角。结合在 mRNA 上的黑点是核糖体。许多核糖体（74 个）的出现是称之为多聚核糖体的原因。随着核糖体接近 mRNA 末端，新生多肽从每个核糖体上延伸开，并变长。新生多肽上弱的小斑点不是单个氨基酸，而是包含氨基酸的结构域。（*Source*：Francke et al.，Electron microscopic visualization of a discreet class of giant translation units in salivary glands of *Chironomus tetans*. *EMBO Journal* 1 1982，pp. 59-62. European Molecular Biology Organization.）

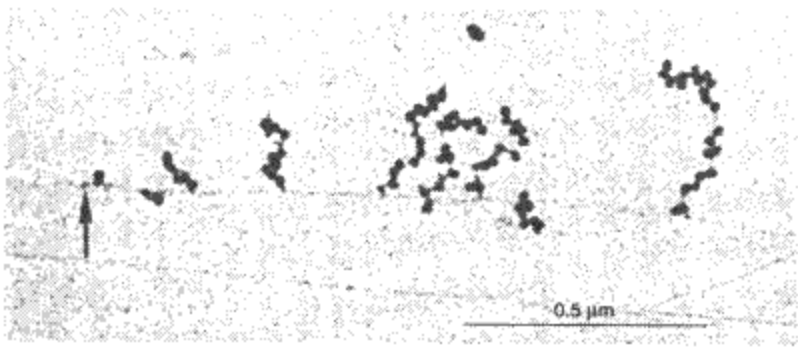


图 19.22　大肠杆菌中转录和翻译的同时发生。两个 DNA 片段水平展开穿过画面，上面的片段从左到右转录。随着 mRNA 增长，越来越多的核糖体结合并进行翻译，由此产生多聚核糖体，它们或多或少地垂直于 DNA 排列。图中看不到新生多肽，左边箭头指示一个弱斑点，可能是刚开始基因转录的 RNA 聚合酶，其他类似的表示 RNA 聚合酶的斑点出现在有些多核糖体的基部，此处 mRNA 结合 DNA。（*Source*：O. L. Miller，B. A. Hamkalo，and C. A. Thomas Jr.，Visualization of bacterial genes in action. *Science* 169（July 1970）p. 394. Copyright © AAAS.）

小结　大多数 mRNA 被多个核糖体在同一时间翻译，形成许多核糖体成串地翻译一条 mRNA 链的结果，这种结构称为多聚核糖体。在真核生物中，多聚核糖体存在于细胞质中。在原核生物中，基因的转录及其所产生 mRNA 的翻译是同时发生的。因此，一个活性基因会结合有很多的多聚核糖体。

19.2　转运 RNA

1958 年，Francis Crick 推测可能存在一种 RNA 的适配分子，作为 DNA（实际在 mRNA 中）核苷酸链和相应蛋白质氨基酸链之间的中介。Crick 倾向于适配分子含有 2 或 3 个能与密码子核苷酸配对的核苷酸这一想法，虽然当时尚无人知道密码子的本质，甚至不知道 mRNA 的存在。Paul Zamecnik 及同事在一年以前已经发现了转运 RNA，虽然他们当时还不知道转运 RNA 起适配分子的作用。

tRNA 的发现

1957 年，Zamecnik 及同事已建成大鼠的无细胞蛋白质合成系统。该系统的组分之一叫做 pH5 酶组分，该组分含有一些可溶性因子与核糖体共同指导所加入 mRNA 的翻译。在 pH5 酶组分中大部分是蛋白质，但他们发现混合物中也包含一种小 RNA。更有趣的是，这个 RNA 能被氨基酸偶联。为证实这一点，他们将这种小 RNA 与 pH5 酶、ATP 以及 [^{14}C] 亮氨酸混合，图 19.23a 表明，加入混合物中的标记亮氨酸越多，结合到 RNA 上的就越多，而去除 ATP 后反应就不能发生。我们现在知道这个反应是 tRNA 对氨基酸的负载。

Zamecnik 及同事发现这个小分子 RNA 不仅能负载氨基酸，还能将其氨基酸传递给正在合成的蛋白质。他们将负载 [^{14}C] 亮氨酸的 pH5 RNA 与微粒体（microsome，含有核糖体的内质网小体）混合。图 19.23b 显示 pH5 RNA 放射性亮氨酸的丢失与微粒体蛋白质中亮氨酸的获得近乎完美对应，这表明亮氨酰-tRNA 的亮氨酸参入了核糖体上的新生多肽。

小结　tRNA 作为一种独立于核糖体的小分子 RNA 被发现，它能负载氨基酸然后传给正在合成的多肽。

tRNA 的结构

要了解 tRNA 怎样行使功能，就需要知道它的分子结构。尽管 tRNA 分子很小，它们的结构却令人吃惊的复杂，就像蛋白质有初级、二级、三级结构一样，tRNA 也有这些结构。初级结构是 tRNA 中

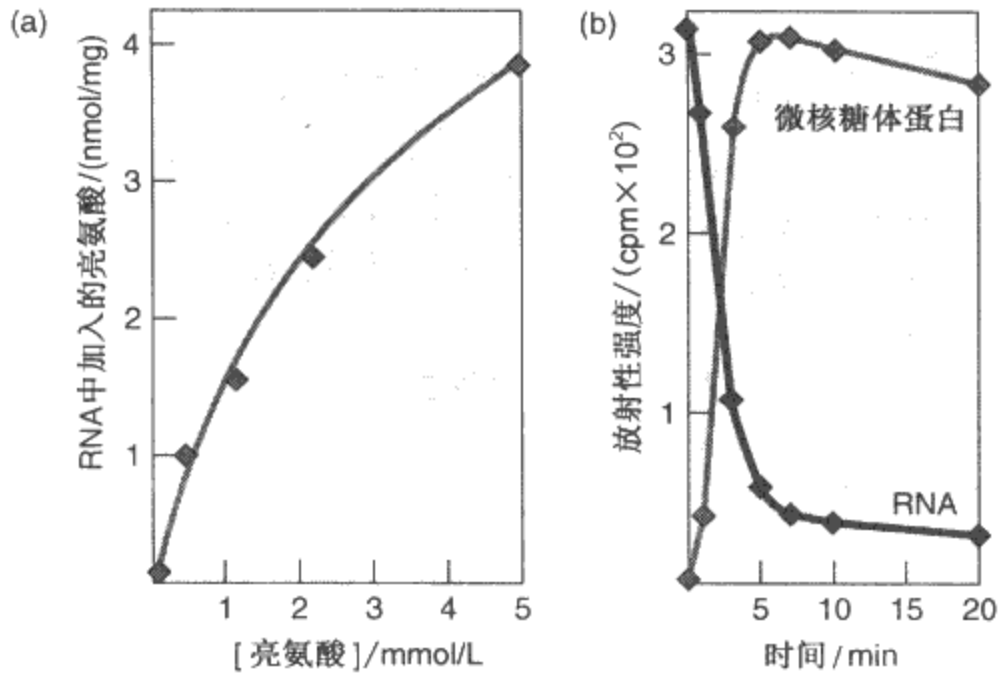


图 19.23　tRNA 的发现。（a）tRNA 可负载亮氨酸。Zamecnik 及同事把标记的亮氨酸加入含有 tRNA 的组分，以亮氨酸对 RNA 的结合量对所加入的标记亮氨酸作图；（b）负载的 tRNA 可以将其氨基酸供给新生蛋白。Zamecnik 及同事记录了 tRNA 放射性（cpm）的丢失（蓝色）和微粒体（包括核糖体）新生蛋白中放射性的获得（红色）。曲线间的互补关系提示 RNA 将其氨基酸贡献给增长的蛋白质。（*Source*：Adapted from Hoagland，M. B.，et al.，*Journal of Biological Chemistry* 231：244 & 252，1958.）

的碱基线性顺序，二级结构是 tRNA 不同区域碱基互相配对形成茎环的方式，三级结构是该分子的整体三维形状。在本节，我们将研究 tRNA 的结构与其功能的关系。

1965 年，Robert Holley 及同事首次测定了酵母丙氨酸 tRNA 天然核酸的碱基序列，其初级序列提示至少存在 3 个令人感兴趣的二级结构，其中包括一个三叶草形状。至 1969 年，已确定了 14 个 tRNA 的序列，结果显示，尽管在初级结构上差别相当大，但所有 tRNA 本质上都具有相同的“三叶草”形二级结构，如图 19.24a 所示。在研究这个结构时，应想到 tRNA 的真实三维结构完全不是三叶草形的，三叶草形只是描述该分子的碱基配对方式而已。

三叶草结构有 4 个碱基对形成的茎，以此确定了分子的 4 个主要区域（图 19.24b）。第一个区域在图的顶部，是 tRNA 的两个末端通过碱基配对形成的**受体茎**（acceptor stem）。3′端具有不变的 CCA 序列，比 5′端突出一些。左边的是**二氢尿嘧啶环**（dihydrouracil loop，D 环），以该部位总是存在修饰过的尿嘧啶命名。底部是**反密码子环**（anticodon loop），因最重要的反密码子在其顶端而得名。如我们在第 3 章所学的，反密码子和 mRNA 密码子碱基配对，因此可对 mRNA 解码。右边是 **T 环**（T loop），因几乎不变的三碱基序列 TΨC 而命名。**Ψ** 代表 tRNA 中一个修饰的核苷**假尿嘧啶核苷**（pseudouridine），除了碱基是通过 5-碳而不是 1-氮连接核糖外，其他与正常尿苷一样。图 19.24 中反密码子环与 T 环之间的区域称为**可变环**（variable loop），其长度变化为 4～13nt，有些较长的可变环含有碱基配对的茎。

除了二氢尿嘧啶和假尿嘧啶外，tRNA 还含有许多修饰的核苷，有些修饰只是简单的甲基化，有些则更精细，如鸟嘌呤核苷转换为 **Y 核苷**（wyosine），含有复杂的称为 Y 碱基的三环碱基（图 19.25）。有些 tRNA 的修饰是一般性的，如所有 tRNA在 T 环的相同位置都有一个假尿嘧啶，多数 tRNA都有一个高度修饰的核苷，如反密码子附近的 Y 核苷。其他修饰是对某些 tRNA 特异性的。图 19.25 给出了 tRNA 中的常见修饰核苷。

tRNA 核苷的修饰产生了一个问题：tRNA 是由修饰碱基合成的，还是转录后完成碱基修饰的？答案是 tRNA 和其他 RNA 的产生方式一样，都由 4 种标准碱基组成，一旦转录完成，多种酶系统就修饰碱基。那么，这些修饰对 tRNA 的功能有什么影响呢？至少两种 tRNA 已经在体外用 4 种正常的、没有修饰的碱基合成了，但它们不能结合氨基酸。因此至少在这两个例子中，完全无修饰的 tRNA 没有功能。虽然这些研究提示所有修饰作用的总和是至关重要的，但是单碱基修饰可能对 tRNA 的负载和利用效率有更细微的影响。

Alexander Rich 及同事在 20 世纪 70 年代用 X 射线衍射技术揭示了 tRNA 的三级结构。因为所有 tRNA 都具有以三叶草模型为代表的本质上相同的二级结构，所以它们本质上具有相同的三级结构也就不足为怪了。图 19.26 显示了酵母 $tRNA^{Phe}$ 分子的倒 L 形结构，该结构最重要的方面可能是将其碱基配对茎的长度最大化，以两个碱基配对茎为一组进行堆积，形成相对伸长的碱基配对区。其中一个配对区水平地位于分子的顶部，包括受体茎和 T 茎，另一个形成分子的垂直轴，包括 D 茎和反密码子茎。由于每个茎的两部分不能完美地排成行而有轻微弯曲，线形排列允许碱基对互相堆积从而赋予其稳定性。该分子的碱基配对茎是 RNA-RNA 双螺旋，我们在第 2 章学过，这种 RNA 螺旋应该呈现 A 螺旋形态，每转一螺旋有 11 个碱基，X 射线衍射研究证实了这一预测。

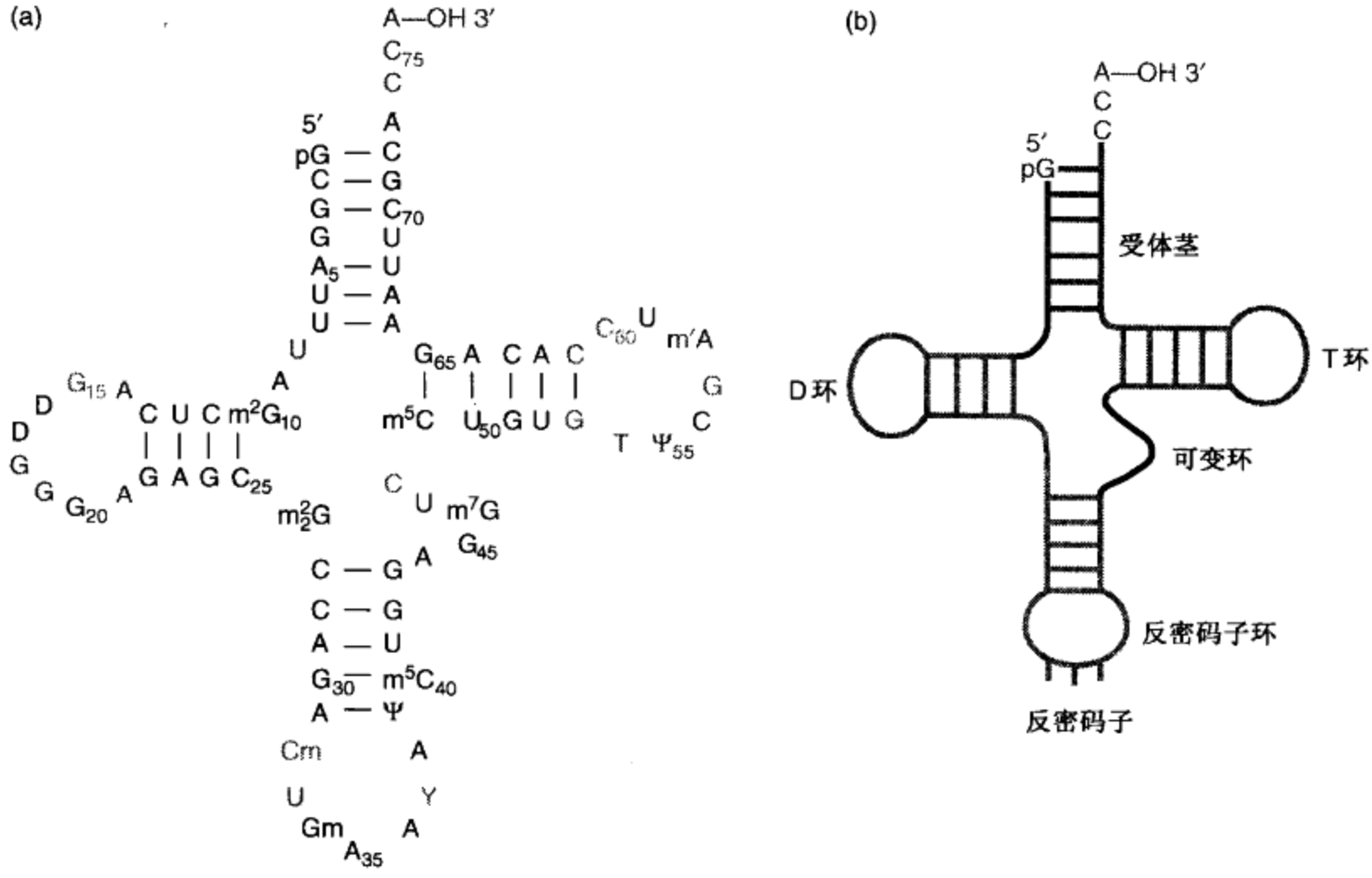


图 19.24　tRNA 三叶草结构的两种视图。（a）酵母 $tRNA^{Phe}$ 的碱基序列，以三叶草形显示，不变核苷酸用红色表示，通常是嘌呤或是嘧啶的碱基用蓝色表示；（b）酵母 $tRNA^{Phe}$ 的三叶草结构。顶部是受体茎（红色），氨基酸和 3′端腺苷结合。左边是二氢尿嘧啶环（D 环，蓝色），至少含有一个二氢尿嘧啶碱基。底部是反密码子环（绿色），含有反密码子。右边是 T 环（灰色），含有不变的 TΨC 序列。每个环与其碱基配对的茎特色。（*Source*：（a）Adapted from Kim, S. H., F. L. Suddath, G. J. Quigley, A. McPherson, J. L. Sussman, A. H. J. Wang, N. C. Seeman, and A. Rich, Three-dimensional tertiary structure of yeast phenylalanine transfer RNA, *Science* 185: 435, 1974.）

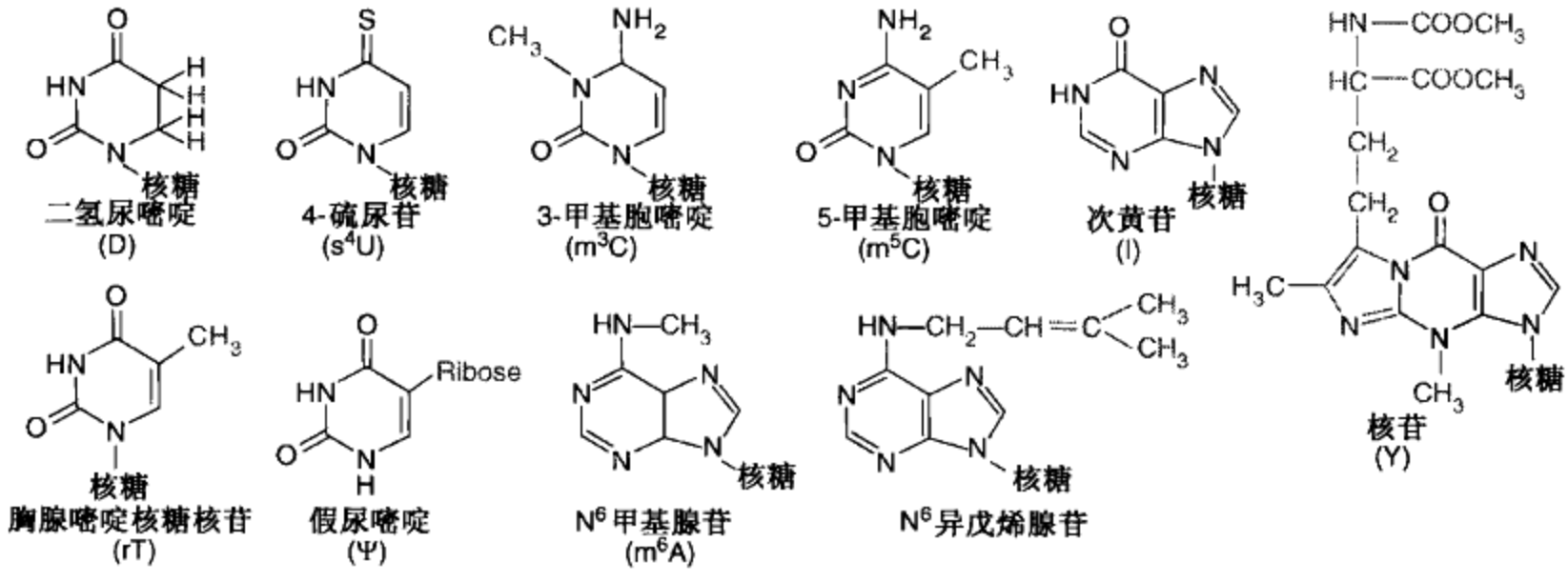


图 19.25　tRNA 中的一些修饰核苷。红色表示对 4 个正常 RNA 核苷中的一个变异。次黄苷是个特例，它是鸟苷和腺苷的前体。

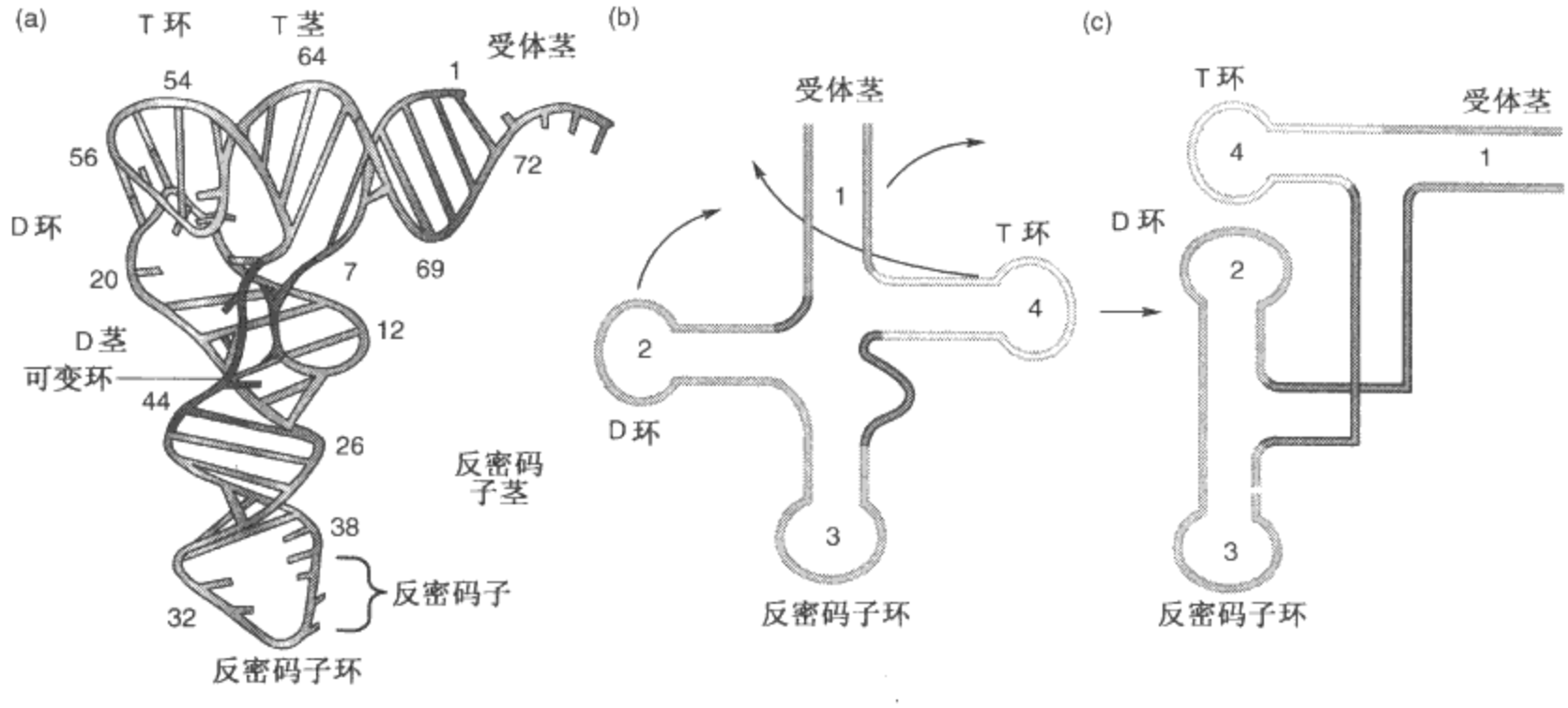


图 19.26　tRNA 的三维结构。（a）酵母 $tRNA^{Phe}$ 三级结构平面投影图。分子各个部分的颜色含义与（b）和（c）对应。（b）熟悉的 tRNA 三叶苜蓿形结构，颜色与（a）相同。箭头指示三叶苜蓿形结构扭曲获得如（c）图所示的接近真实 tRNA 形状的方向。（*Source*：Adapted from Quigley，G. J. and A. Rich，Structural domains of transfer RNA molecules，*Science* 194：197 Fig. 1b，1976.）

图 19.27 是酵母 $tRNA^{Phe}$ 分子的立体图，碱基配对区在三维图中很容易看到，但在 T 茎-受体区甚至可在二维结构上观察到，因为它们被画成几乎垂直于纸面，以平行线呈现。

正如已经看到的，tRNA 主要是通过形成碱基配对区的二级相互作用而稳定，也可通过区域间的几十个三级相互作用达到稳定，包括碱基-碱基、碱基-骨架、骨架-骨架相互作用。多数包括氢键的碱基-碱基三级互作发生在不变碱基或半不变碱基之间（半不变碱基总是嘌呤碱或者嘧啶碱）。这些相互作用使 tRNA 能正确折叠，所以相关碱基应保持不变，任何变化都会影响正确折叠进而阻碍 tRNA 正确行使功能。

只有一种碱基-碱基相互作用是正常的 Watson-Crick 碱基配对的，其他所有都是反常的。例如，使 D 环与可变环连接的 G15-C48 碱基对，两条链是平行的而不是反向平行的，所以不是 Watson-Crick 配对，称为**反式配对**（*trans*-pair）。还有几个例子是一个碱基与另外两个碱基相互作用，其中有 U8、A14 和 A21。我们已讨论过三级互作，再看一下图 19.26a，可观察其更真实的形态。例如，碱基 18 和 55 间以及碱基 19 和 56 间的相互作用。乍一看，好像是 T 环内的碱基对，仔细观察发现它们连接着 T 环和 D 环。

tRNA 三级结构另一个值得注意的方面是反密码子结构。图 19.27 证明反密码子碱基是堆积的，但这种堆积随着碱基向右突出，远离 tRNA 骨架，使它们处于和 mRNA 密码子碱基相结合的位置。实际上，反密码子骨架已被扭成部分螺旋状，这可能促进了与相应密码子的碱基配对（图 19.2）。

小结　所有 tRNA 都具有共同的表现为三叶草形的二级结构，有 4 个碱基配对的茎形成 3 个茎环结构（D 环、反密码子环和 T 环）和一个受体茎，氨基酸在负载步骤被加到受体茎上。tRNA 也具有共同的三维形状，类似一个倒置的 L。这种形状通过将 D 茎与反密码子茎的碱基对以及 T 茎与受体茎的碱基对线性排列起来而使稳定性最大。tRNA 的反密码子从反密码子环的侧面突出，并被扭成一种与 mRNA 中相应的密码子容易配对的形状。

氨酰-tRNA 合成酶对 tRNA 的识别：第二遗传密码

1962 年，Fritz Lipmann 及同事证明，核糖体识别氨酰-tRNA 中的 tRNA 而不是氨基酸。他们生成了半胱氨酰-$tRNA^{Cys}$，然后用兰尼镍还原半胱氨酸产生丙氨酰-$tRNA^{Cys}$，如图 19.28 所示［注意命名，在半胱氨酰-$tRNA^{Cys}$（Cys-$tRNA^{cys}$）中，第一个 Cys 表明实际结合在 tRNA 上的氨基酸，第二个 Cys（右上标）表示这个 tRNA 应该结合的氨基酸。因此，丙氨酰-$tRNA^{Cys}$ 应该是结合半胱氨酸的 tRNA，但却结合了丙氨酸］。然后，按 5∶1 的比例将这种改变的氨酰-tRNA 和人工合成的 mRNA 加入体外翻译系统。该 mRNA 是 U 和 G 的随机多聚体，

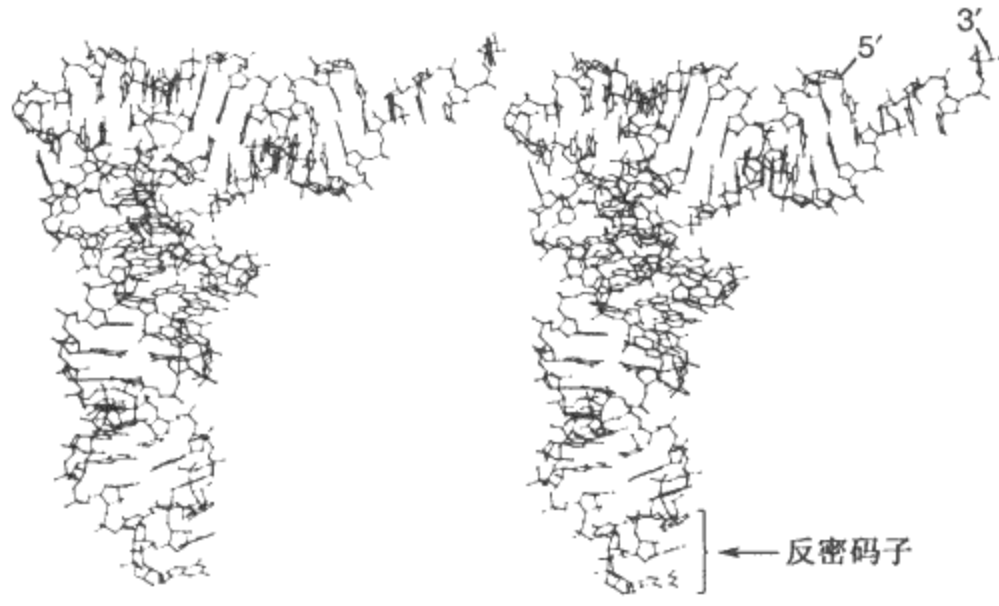


图 19.27　tRNA 的立体结构图。要从三维角度观察，需要利用立体阅读器，或使两个图像相融合，方法是：放松眼睛，好像在这个距离上聚焦于某物（即“魔术眼”技术），或是微闭眼睛。产生三维立体效果可能要花一点时间。（*Source*：Adapted from Quigley，G. J and A. Rich，Structural domains of transfer RNA molecules. *Science* 194（19 Nov 1976）f. 2，p. 798.）

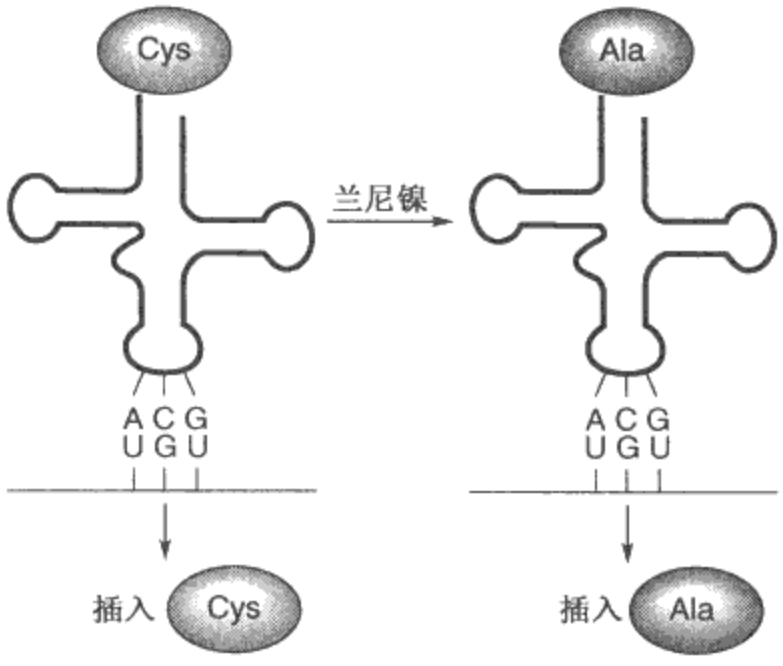


图 19.28　核糖体识别氨酰-tRNA 中的 tRNA 而不是其氨基酸。Lipmann、Ehrenstein、Benzer 及同事从 cysteyl-$tRNA^{Cys}$ 着手，该氨酰-tRNA 将半胱氨酸插入蛋白质链中，如图左边所示。用兰尼镍处理该氨酰-tRNA，使半胱氨酸还原为丙氨酸，但对 tRNA 无影响。该丙氨酰-$tRNA^{Cys}$ 将丙氨酸插入蛋白质链中通常是半胱氨酸的位置，如图右边所示。因此，连接到tRNA上的氨基酸性质并不重要，重要的是 tRNA 自身，因为它的反密码子必须与 mRNA 的密码子配对。

含多个编码半胱氨酸的 UGU 密码子，可正常性地引起半胱氨酸的加入，但不应该引起丙氨酸的加入，因为丙氨酸的密码子是 GCN，这里 N 是任意碱基，而 UG 多聚体不含有 C。但是在本例中，丙氨酸因结合了 $tRNA^{Cys}$ 而被加入，这表明核糖体不区别结合到 tRNA 上的氨基酸，它们只识别氨酰-tRNA 的 tRNA 部分。

这个实验指出氨酰-tRNA 合成酶忠实性的重要性。核糖体只识别氨酰-tRNA 的 tRNA 部分这一事实，意味着如果合成酶出错并将错误的氨基酸放到 tRNA 上，那么这些氨基酸将在错误的位置被插入蛋白质中。这是非常危险的，因为含有错误氨基酸序列的蛋白质很可能无法正确行使功能。因此，氨酰-tRNA 合成酶对 tRNA 及其携带的氨基酸非常专一就不足为奇了。由此产生了关于 tRNA 结构的主要问题：假设所有 tRNA 的二级和三级结构本质上是一样的，那么当合成酶从 tRNA 库的二十多种中选出一种 tRNA 时是以什么碱基序列进行识别呢？这组序列已被命名为“第二遗传密码”以突出其重要性。该问题的复杂性在于 tRNA 的**同工种类**（isoaccepting species）可由同一合成酶加载同样的氨基酸，然而它们却有不同的序列甚至不同的反密码子。

如果推测氨酰-tRNA 合成酶识别 tRNA 元件的部位，我们可能会想到两个位点。第一，受体茎是一个合理的选择，因为它是 tRNA 上接受氨基酸的位点，因此有可能在负载氨基酸时位于或接近酶活性中心。推测酶和受体茎有非常密切的接触，所以它应该能识别受体茎上具有不同碱基顺序的 tRNA。当然，最后 3 个碱基都是相同的 CCA 与这个目的无关。第二，反密码子是一个合理的选择，因为它在每个 tRNA 中都不相同，并且与 tRNA 应该加载的氨基酸有直接关系。我们将看到多数情况下这两种推测都是对的，并且某些 tRNA 的其他部位在氨酰-tRNA 合成酶识别过程中也起作用。

受体茎　1972 年，Dieter Söll 及同事注意到在多数 tRNA 的 73 位，即 3′端起第 4 个碱基的性质有一种模式：该碱基在专对某一类氨基酸的 tRNA 中趋于相同。例如，无论是何种 tRNA，几乎所有亲水性氨基酸都通过 73 位的 A 与 tRNA 结合。这显然不是故事的全部，因为一个碱基不可能提供足够的变化来解释 20 种不同 tRNA 的专一性负载，它最多是作为一个初步的识别因子。

Bruce Roe 和 Bernard Dudock 用了另一种方法。

他们分析了能被单个合成酶加载的所有各种 tRNA 的碱基序列，包括在异种错误加载（heterologous mischarging）过程中由错误氨基酸加载的一些 tRNA。异种错误加载是指一种合成酶对另一种不正确 tRNA 进行加载的能力，这种错误加载比较慢，且需要比正常加载高的酶浓度。例如，酵母苯丙氨酰-tRNA 合成酶（PheRS）可以正确地加载大肠杆菌、酵母以及小麦胚芽的 $tRNA^{Phe}$，也能用苯丙氨酸加载大肠杆菌的 $tRNA^{Val}$。

因为所有这些 tRNA 都能被相同的合成酶加载，所以它们应该都有一个元件让合成酶来分辨加载哪个 tRNA。Roe 和 Dudock 比较了所有这些 tRNA 的序列，寻找它们所共同有的东西，但所有 tRNA 没有共同之处。但有两点凸显出来：73 位碱基和 D 茎的 9 个核苷酸。

1973 年，J. D. Smith 和 Julio Celis 研究了一种突变体抑制剂 tRNA，它可插入 Tyr 而不是 Gln。换句话说，野生型抑制子 tRNA 被 Gln RS 所加载，但其序列上发生某个改变后可被 TyrRS 所加载，突变体和野生型的唯一不同是 73 位的碱基由 G 变为 A。

1988 年，Ya-Ming Hou 和 Paul Schimmel 用遗传学方法证明受体茎的单碱基对对其加载专一性很重要。他们将 $tRNA^{Ala}$ 的反密码子突变为 5′-CUA-3′，变成琥珀抑制子以便对应琥珀密码子 UAG 插入丙氨酸，然后寻找在 tRNA 中改变其加载特异性的突变。分析方法很简便，可在体内进行。他们构建了一个在密码子 10 位具有琥珀突变的 *trpA* 基因，该突变只能被响应琥珀密码子插入丙氨酸（或甘氨酸）的 tRNA 所抑制。10 位上的任何其他氨基酸都只产生无活性的蛋白质。最后，在缺乏色氨酸条件下培养突变体，如果突变体能抑制 *trp*A 基因的琥珀突变，那么它可能含有能被丙氨酸加载的抑制子 tRNA，如果不能，则表明抑制子 tRNA 被改变了，可被另一氨基酸加载。他们发现，在缺乏色氨酸环境中生长的所有细胞在抑制子 tRNA 的 3 位有一个 G，70 位有一个 U，所以在受体茎距末端 3 个碱基处能形成 G3-U70 摇摆碱基对。

这个实验提示，G3-U70 碱基对是 Ala RS 加载的重要决定因素。如果是这样，研究者推测可以插入另一个氨基酸的另一个抑制子 tRNA，通过将其 3 位和 70 位碱基分别变为 G 和 U，使其加载专一性转换为丙氨酸。他们采用了两种不同的抑制子 tRNA：$tRNA^{Cys/CUA}$ 和 $tRNA^{Phe/CUA}$，这里的 CUA 指识别琥珀密码子 UAG 的反密码子。这两个 tRNA 的受体茎上都有一个 C3-G70 碱基对。当 Hou 和 Schimmel 将这个碱基对变为 G3-U70 时，就将 tRNA 转变为 $tRNA^{Ala/CUA}$，以它们能抑制在 *trp*A 基因密码子 10 上的琥珀突变为指示。

这些改变的琥珀抑制子 tRNA 真的能将丙氨酸插入 TrpA 蛋白吗？氨基酸序列测定揭示确实如此！而且这些改变的 tRNA 在体外能够被丙氨酸加载。因此，即使这两种 tRNA 分别在 38 和 31 位碱基不同于天然的 $tRNA^{Ala/CUA}$，只将一个碱基对从 C-G 变为 G-U 就可将其加载专一性分别从 Cys 和 Phe 改变为 Ala。

1989 年，Christopher Francklyn 和 Schimmel 提供了另一个证据表明受体茎，具体讲是 G3-U70 碱基对，在 AlaRS 负载专一性方面的重要性。他们发现，用人工合成的 35nt “微螺旋”模拟 L 型 $tRNA^{Ala}$ 的上部，包括受体茎和 TΨC 环，可以有效地负载丙氨酸。实际上，只要存在 G3-U70 碱基对，即使有很多碱基被改变，对丙氨酸的负载也可以发生。

同样很有意思的是，Ala-微螺旋可结合到核糖体的 P 位点，像完好的 Ala-$tRNA^{Ala}$ 一样参与肽基转移酶与嘌呤霉素的反应。根据这些观察可以推测，tRNA 分子的顶部首先进化，与 23S rRNA 祖先一起参与核糖体进化之前“RNA 世界”原始版本的蛋白质合成。

小结 生物化学和遗传学实验已经证明受体茎在同类氨酰-tRNA 合成酶对 tRNA 识别过程中的重要性。在某些情况下，改变受体茎的一个碱基对可以改变加载特异性。

反密码子 1973 年，LaDonne Schulman 开发了一项技术，用亚硫酸氢盐处理 $tRNA_f^{Met}$，使胞嘧啶转变为尿嘧啶。她与同事发现，多数情况下，这种碱基改变都没有效果，但是有些破坏了 tRNA 对甲硫氨酸负载的能力。其中一种改变是 73 位碱基 C→U，另一个是反密码子的 C→U。从那以后，Schulman 及同事收集了大量证据表明反密码子在加载专一性方面的重要性。

1983 年，Schulman 和 Heike Pelka 建立了一种方法，可在初始 tRNA（$tRNA_f^{Met}$）的反密码子中特异性地一次改变一个或多个碱基。首先，用胰腺 RNase 进行限制性消化，将野生型 tRNA 切为两部分：5′片段去掉反密码子，3′片段切除 CCA 末端的最后两个核苷酸。然后用 T4RNA 连接酶将一个小的寡核苷酸片段连接到 5′片段上替代失去的反密码子，但有一个或多个碱基改变。然后将这两部分重新连到一起，并且用 tRNA 核苷酸转移酶重新加上失去的 CA 尾。最后，他们在体外加载反应中检测改变了反密码子的 tRNA。表 19.1 显示，改变 $tRNA_f^{Met}$ 反密码子的一个碱基就足以使 Met 的加载率降低为原先的 $1/10^5$。反密码子的第一个碱基

（摇摆位置）是最敏感的，改变这个碱基总是对加载有显著的影响。因此，反密码子可能是该 tRNA 在体外加载所必需的。

表 19.1　$tRNA_f^{Met}$ 衍生物氨酰化的初始速率

tRNA*	Met-tRNA 摩尔数/Met-tRNA 合成酶摩尔数每分钟	相对速率，CAU/其他
$tRNA_f^{Met}$	28.45	0.8
$tRNA_f^{Met}$(gel)†	22.80	1
CAU	22.15	1
CAUA	1.59	14
CCU	4.0×10^{-1}	55
CUU	2.6×10^{-2}	850
CUA	2.0×10^{-2}	1 100
CAG	1.7×10^{-2}	1 300
CAC	1.2×10^{-3}	18 500
CA	0.5×10^{-3}	44 000
C	$<10^{-4}$	$>10^5$
ACU	$<10^{-4}$	$>10^5$
UAU	$<10^{-4}$	$>10^5$
AAU	$<10^{-4}$	$>10^5$
GAU	$<10^{-4}$	$>10^5$

* 插入人工合成的 $tRNA_f^{Met}$ 衍生物的反密码子环中的寡核苷酸。

† 从变性聚丙烯酰胺凝胶中分离的对照样品，平行样品是人工合成的 $tRNA_f^{Met}$ 衍生物。

Source: L. H. Schulman and H. Pelka, "Anticodon Loop Size and Sequence Requirements for Recognition of Formylmehionine tRNA by Methionyl-tRNA Synthetase," *Proceedings of the National Academy of Sciences* November 1983. Reprinted with permission of the author.

1991 年，Schulman 和 Leo Pallanck 根据以前的体外研究，在体内研究了反密码子改变的影响。同样，他们改变了 $tRNA_f^{Met}$ 的反密码子，但是这次他们检测改变了的 tRNA 错误加载对应新的反密码子氨基酸的能力。他们用编码二氢叶酸还原酶（dihydrofolate reductase，DHFR）的报告基因来检测错误加载，高纯度的该基因产物很容易分离。以下是该分析如何进行的例子：改变编码 $tRNA_f^{Met}$ 的基因，其反密码子由 CAU 变为 GAU，这是异亮氨酸（Ile）反密码子。然后，将这个突变基因连同携带 AUC 起始密码子的突变型 DHFR 基因导入大肠杆菌。

通常情况下，AUC 不能像起始密码子那样很好地工作，但是在带有互补反密码子的 $tRNA_f^{Met}$ 存在时它可以工作，如表 19.2 所示。对所得 DHFR 蛋白质序列的分析表明，第一个位置上的氨基酸主要是 Ile，也有一些甲硫氨酸，表明内源野生型 $tRNA_f^{Met}$ 在一定程度上能够识别 AUC 初始密码子。

Pallanck 和 Schulman 用同样的方法将 $tRNA_f^{Met}$ 的反密码子改变为 GUC（缬氨酸，Val）或 UUC（苯丙氨酸，Phe）。在每种情况下，他们都对 DHFR 的起始密码子做相应改变，使其与改变的 $tRNA_f^{Met}$ 中的反密码子互补。表 19.2 证明在这两种情况下，基因功能在有互补 $tRNA_f^{Met}$ 时明显好于缺乏互补 $tRNA_f^{Met}$ 时。更重要的是，起始氨基酸的性质可随 tRNA反密码子的改变而改变。实际上，存在携带缬氨酸反密码子的 $tRNA_f^{Met}$ 时，缬氨酸是唯一在 DHFR 蛋白质氨基端发现的氨基酸，这表明 $tRNA_f^{Met}$ 反密码子由 CAU 到 GAC 的变化将这个 tRNA 的加载特异性从甲硫氨酸改变为缬氨酸。因此，在这个例子中，反密码子好像是决定 tRNA 加载特异性的重要因素。

表 19.2　在有或无互补的 $tRNA_f^{Met}$ 反密码子突变体条件下 DHFR 在甲硫氨酸和非甲硫氨酸起始密码子处的合成

		DHFR，单位			
起始密码子	$tRNA_f^{Met}$ 反密码子	＋互补 $tRNA_f^{Met}$	－互补 $tRNA_f^{Met}$	相对于 AUG 的效率/%	NH_2端氨基酸/%
5′-AUG-3′	5′-CAU-3′	1085±94	1109±133	100	Met(100)
AUC	GAU	298±26	104±10	23	Ile(84), Met(16)
GUC	GAC	201±3	2.7±0.1	18	Val(100)
UUC	GAA	61±11	2.0±0.5	4.3	Phe(76), Met(21), Ile(3)
UGC	GCA	5.0±0.3	0.5±0.08	0.4	—
AAC	GUU	4.1±0.1	0.5±0.04	0.3	—
ACC	GGU	3.8±0.4	0.5±0.05	0.3	—
UUA	UAA	3.4±0.3	1.2±0.2	0.2	—
GGC	GCC	2.8±0.2	0.5±0.04	0.2	—
GAC	GUC	0.7±0.3	0.4±0.03	0.03	—

Source: From "Anticodon-Dependent Aminoacylation of a Noncognate tRNA with isoleucine, Valine, and Phenylalanine in Vivo, *Proceedings of the National Academy of Sciences*, 88: 3872-76, May 1991. Reprinted with permission of the author.

此外，需注意改变 $tRNA_f^{Met}$ 的反密码子总是使它的效率降低。实际上，在表 19.2 底部列出的改变所产生的 $tRNA_f^{Met}$ 分子的效率太低了，甚至在有互补起始密码子时也无法进一步分析。因此，有的氨酰-tRNA 合成酶能加载带有改变的反密码子的非同类 tRNA，有的则不行。对后面的这些酶显然需要比反密码子更多的信息。

小结 生物化学和遗传学实验已经表明反密码子和受体茎一样，是负载专一性的一个重要元素。有时候反密码子能够完全决定负载的特异性。

合成酶-tRNA 复合体的结构 对 tRNA 及其同类氨酰-tRNA 合成酶复合体的 X 射线晶体学研究表明，受体茎和反密码子在合成酶上都有锚定位点。这些发现强调了受体茎和反密码子在合成酶识别中的重要性。1989 年，Dieter Söll 及同事用 X 射线晶体学技术确定了第一个氨酰-tRNA 合成酶（*E. coli* GInRS）结合其关联 tRNA 的三维结构。图 19.29 显示了这一结构。靠近顶部，可以看到酶上有一条深的裂缝把受体茎包进去，包括碱基 73 和3～70 碱基对。在左下部，可观察到酶上有一条较小的裂缝，tRNA 的反密码子伸入其中。这样允许反密码子被合成酶特异性识别。另外，我们看到酶的最左侧的大部分区域与 tRNA 的 L 形的内侧紧密接触，包括 D 环侧面和受体茎的小沟。

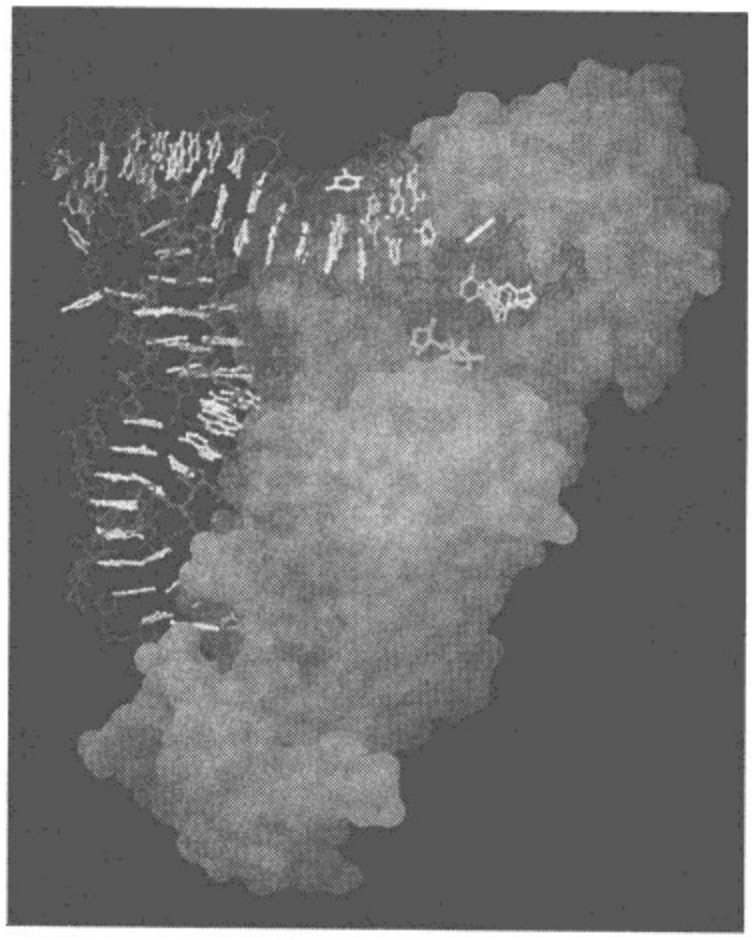


图 19.29 含有 tRNA 和 ATP 的谷氨酰胺-tRNA 合成酶的三维结构。合成酶为蓝色，tRNA 为棕或黄色，ATP 为绿色。注意在酶与 tRNA 之间的 3 个接触区。①包被 tRNA 受体茎及 ATP 的顶部深沟；②左下方容纳 tRNA 反密码子的口袋；③两个深沟间的区域与 tRNA 的 L 形内侧大部分区域接触。（*Source*：Courtesy T. A. Steitz; from Rould, Perona, Vogt, and Steitz, *Science* 246（1 Dec 1989）cover. Copyright © AAAS.）

约有一半的合成酶包括 GlnRS 可归为一类，叫做 **class I**，它们在结构上相似，并且最先氨酰化 tRNA腺苷末端的 2′-羟基。另一半合成酶为 **class II**，它们在结构上彼此相似，但十分不同于 class I，它们最先氨酰化关联 tRNA 的 3′-羟基。1991 年，D. Moras 及同事获得了该组成员的酵母 AspRS 连同 $tRNA^{Asp}$ 的 X 射线晶体结构。图 19.30 对比了 class I 和 class II 合成酶-tRNA 复合体的结构。有几个不同之处，首先，class II 合成酶虽然仍然接触 L 结构的内侧，但同样也接触 tRNA 的另一面，包括可变环和受体茎的大沟。另外，受体茎包括 CCA 尾，是一种规则的螺旋构象。这与 class I 结构相反，在 class I 中第一个碱基对被破坏，分子的 3′端形成发夹结构。因此，X 射线晶体学研究证实了对合成酶-tRNA相互作用的生物化学和遗传学研究的主要结论：反密码子和受体茎都与酶密切接触，因此处于决定酶-tRNA 相互作用特异性的位置。

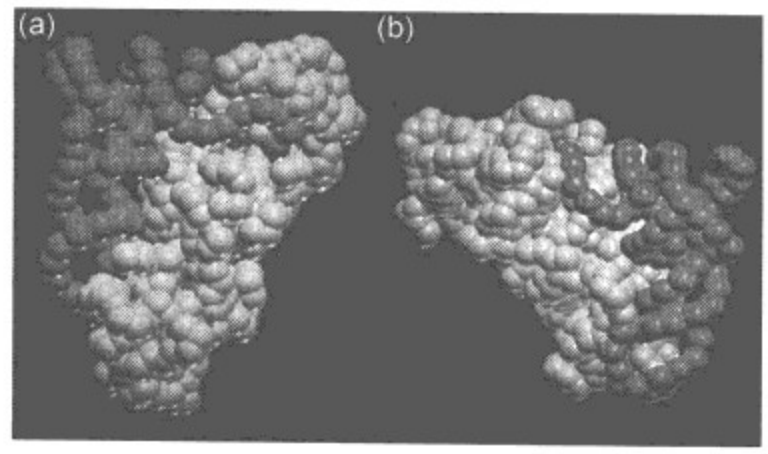


图 19.30 class I 复合体 GInRS-$tRNA^{Gln}$（a）和 class II 复合体酵母 AspRS-$tRNA^{Asp}$（b）的模型。为简洁起见，只显示 tRNA 磷酸盐骨架（红色）及合成酶的 α-C 骨架（蓝色）。注意，两个合成酶对其关联 tRNA 的接近从相反的两面进行。（*Source*：Ruff, M., S. Krishnaswamy, M. Boeglin, A. Poterszman, A. Mitschler, A. Podjarny, B. Rees, J. C. Thierry, and D. Moras, Class II aminoacyl transfer RNA synthetases: Crystal structure of yeast aspartyl-tRNA synthetase complexed with $tRNA^{Asp}$. *Science* 252（21 June 1991）f. 3, p. 1686. Copyright © AAAS.）

小结 X 射线晶体结构表明在两组氨酰-tRNA合成酶之间，合成酶-tRNA 的相互作用是不

同的。class I合成酶有针对受体茎和关联 tRNA 反密码子的口袋，从 D 环和受体茎小沟一侧靠近 tRNA。class II 合成酶也有针对受体茎和反密码子的口袋，但是从另一面包括可变臂和受体茎的大沟接近tRNA。

氨酰-tRNA 合成酶的校正和编辑

正如氨酰-tRNA 合成酶正确识别（关联）tRNA一样，另一个更困难的工作是识别关联氨基酸。原因很明显，tRNA 是大的复杂分子，在核酸序列以及核苷酸修饰上不同分子都有变化，而氨基酸是简单分子，不同分子间彼此相似，有时非常相似。例如，异亮氨酸和缬氨酸，除了异亮氨酸多一个甲基（CH_2）外，其他部分完全一样。1958 年，Linus Pauling 用热动力学方法计算出异亮氨酰-tRNA合成酶（IleRS）应该像正确 Ile-$tRNA^{Ile}$ 配对一样产生约 1/5 的错误 Val-$tRNA^{Ile}$ 配对。而实际上，在 150 个被 IleRS 激活的氨基酸中仅有一个是缬氨酸，并且由该酶产生的 3000 个氨酰-tRNA 中仅有一个是 Val-$tRNA^{Ile}$。异亮氨酰-tRNA 合成酶是怎样阻止 Val-$tRNA^{Ile}$ 的形成呢?

1977 年 Alan Fersht 首先推测，合成酶用一种双筛（double-sieve）机制避免产生带有错误氨基酸的 tRNA。图 19.31 阐释了这一概念。第一次筛选由酶的**活化位点**（activation site）来完成，该位点拒绝太大的底物，而像缬氨酸这样的小底物恰好能进入活化位点并且活化为氨酰腺苷酸形式，有时直接成为氨酰-tRNA 形式。然后第二次筛选发挥作用，活化的氨基酸，或少数情况下太小的氨酰-tRNA 被酶的另一个**编辑位点**（editing site）水解。

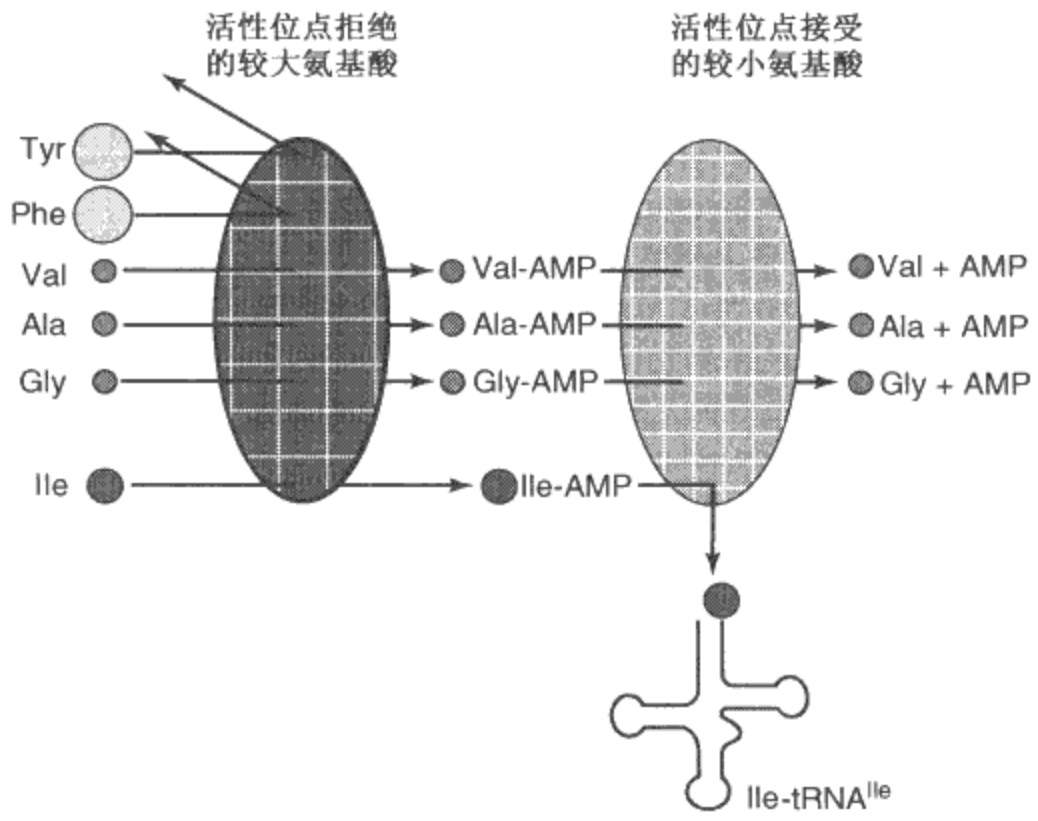


图 19.31　异亮氨酸-tRNA 合成酶的双筛选。活化位点是粗筛，大氨基酸如 Tyr 和 Phe 因为不适合而被排除在外。编辑（水解）位点是细筛，接受比 Ile-AMP 小的活化氨基酸，如 Val-AMP、Ala-AMP、Gly-AMP，但是拒绝 Ile-AMP，因其太大。结果，较小的活化氨基酸被水解为 AMP 和氨基酸，而 Ile-AMP 被转化为 Ile-$tRNA^{Ile}$。（左图：活化位点——排除较大的氨基酸；右图：编辑位点——接受较小的氨酰-AMP）（*Source*: Adapted from Fersht, A. R., Sieves in sequence. *Science* 280: 541, 1998.）

例如，IleRS 利用第一筛排除过大的或错误形状的氨基酸，苯丙氨酸因其太大而被排斥，亮氨酸因其形状不对（亮氨酸的一个末端甲基不能进入活性位点）而被排斥。但是像缬氨酸这样的小氨基酸会怎样呢? 它们确实能进入 IleRS 的活化位点并且被活化，但随后被转运到编辑位点，在那里被识别为不正确并被降解。这种第二次筛选也被称为**校正**（proofreading）或**编辑**（editing）。

Shigeyuki Yokoyama 及同事已经获得嗜热菌 IleRS 的晶体结构及其偶联有关联氨基酸异亮氨酸、非关联氨基酸缬氨酸的晶体结构，这些结构有力地证明了 Fersht 的假设。图 19.32 显示其活化位点的结构，图 19.32a 是结合有异亮氨酸的活化位点，图 19.32b 是结合有缬氨酸的活化位点。可以看到两个氨基酸都能很好地嵌入活化位点中，虽然缬氨酸与活化位点周围的两个疏水性氨基酸侧链（Pro46 和 Trp558）的接触较弱。此外，很明显这个位点太小，不能容纳像苯丙氨酸这样的大氨基酸，甚至亮氨酸也会由于其两个末端甲基中的一个而被阻止结合。该图与双筛假说的初筛部分完全一致。

该酶还有大小与活化位点的裂缝相当的第二个深裂缝，但是相距 34 Å。这个裂缝被认为是编辑位点，部分原因是基于含有这一裂缝的酶片段仍然保持编辑活性。晶体结构证实了这个假说，当 Yokoyama 及同事制备带有缬氨酸的 IleRS 结晶时，在裂缝的底部发现了一个缬氨酸分子。但是制备带有异亮氨酸的 IleRS 结晶时，在裂缝内没有发现氨基酸。因此，裂缝好像对缬氨酸是特异的，它可能是编辑位点。此外，对发现缬氨酸的“口袋”观察表明，Trp232 和 Tyr386 侧链之间的空间刚好够容纳缬氨酸（图 19.33），但是对接纳异

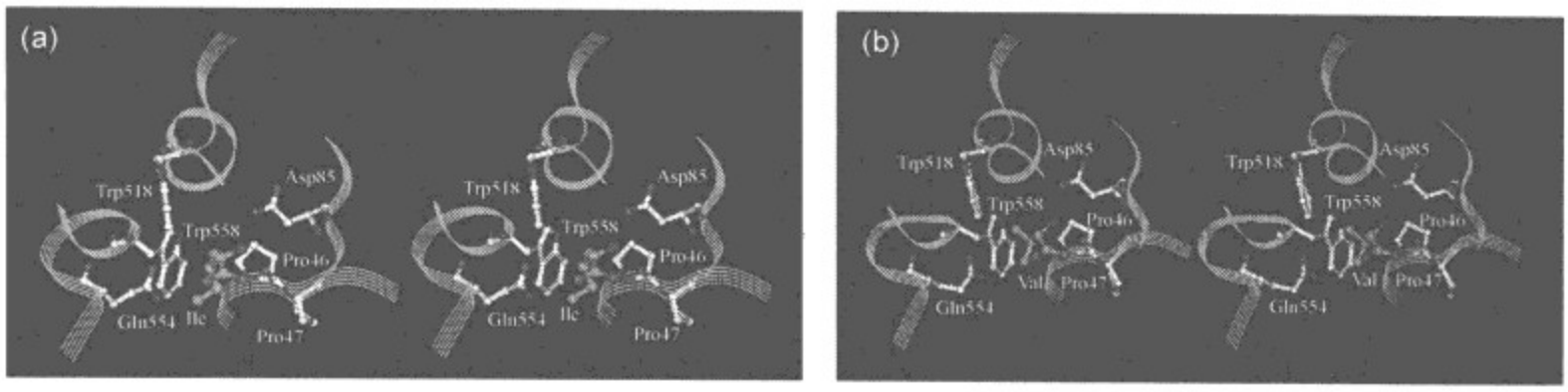


图 19.32　在 IleRS 活性位点的异亮氨酸和缬氨酸立体图。酶的骨架为蓝色带状，氨基酸侧链的碳为黄色。底物的碳原子为绿色［（a）异亮氨酸，（b）缬氨酸］。所有氨基酸的氧原子为红色，氮原子为蓝色。注意异亮氨酸和缬氨酸都进入了活性位点。（*Source*：Nureki，O.，D. G. Vassylyev，M. Tateno，A. Shimada，T. Nakama，S. Fukai，M. Konno，T. L. Henrickson，P. Schimmel，and S. Yokoyama，Enzyme structure with two catalytic sites for double-sieve selection of substrate. *Science* 280（24 Apr 1998）f. 2，p. 579. Copyright © AAAS.）

亮氨酸来讲太小。

如果这里真是编辑位点，那么去除它就可能消除编辑功能。确实，当 Yokoyama 及同事从这一区域移去 47 个氨基酸包括 Trp232 时，编辑活性消失，但保留了全部的活化活性。因此，第二个裂缝确实表现为编辑位点。有几个氨基酸侧链特别靠近裂缝里的缬氨酸，并且 Thr230 和 Asn237 恰好处于参与水解反应的位置，水解反应是编辑的实质。为检测这一假说，Yokoyama 及同事改变了大肠杆菌IleRS中的一些氨基酸（Thr243 和 Asn250），这些氨基酸对应于嗜热细菌酶的 Thr230 和 Asn237。将这些氨基酸变为丙氨酸时，酶失去了编辑活性，但是保留其活化活性。所有这些结果都与第二裂缝是编辑位点这一假说一致，非关联氨酰-AMP 如 Val-AMP 的水解就在这里发生。

小结　至少部分氨酰-tRNA 合成酶的氨基酸选择性由双筛机制控制。第一次筛选是粗筛，排除太大的氨基酸，由合成酶与氨基酸活化的活化位点一起完成这一任务，该位点刚够容纳关联氨基酸，但不能是更大的氨基酸。第二次筛选是细筛，降解太小的氨酰-AMP，由合成酶与接纳小氨酰-AMP 并将其水解的第二活化位点（编辑位点）一起完成这一任务。关联氨酰-AMP 由于过大不能嵌入编辑位点，所以可避免水解，由合成酶将活化的氨基酸转运给它的同类 tRNA。

总结

X 射线结晶学对有或无 tRNA 时细菌核糖体结构的研究已经表明 tRNA 占据两亚基之间的裂缝。它们通过其反密码子末端和 30S 亚基连接，并且通过其受体茎和 50S 亚基连接。核糖体中 tRNA 结合位点主要是由 rRNA 组成的。A、P 位点的tRNA反密码子相互间紧密靠近足以和邻近的结合到 30S 亚基上的密码子碱基配对，两个密码子间的 mRNA 扭转了 45°。在 50S 亚基的肽基转移酶口袋内，A、P 位点上 tRNA 的受体茎也相互紧密靠近，仅相距 5Å。可以看到亚基间有 12 个接触点。

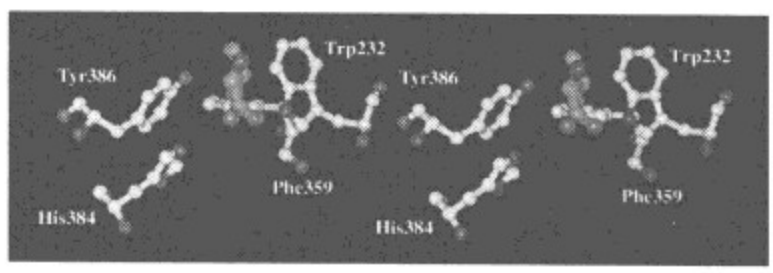


图 19.33　结合到 IleRS 编辑位点的缬氨酸立体图。颜色与图 19.32 相同。（*Source*：Nureki，O.，D. G. Vassylyev，M. Tateno，A. Shimada，T. Nakama，S. Fukai，M. Konno，T. L. Henrickson，P. Schimmel，and S. Yokoyama，Enzyme structure with two catalytic sites for double-sieve selection of substrate. *Science* 280（24 Apr 1998）f. 2，p. 579. Copyright © AAAS.）

*E. coli*核糖体的晶体结构有两种，差异是由核糖体结构域的相对刚体运动引起的。具体讲，30S 颗粒的头部转动了 6°，相比之下，*T. thermophilus* 核糖体转动了 12°。这种转动可能是移位过程中核糖体做棘轮运动的一部分。

大肠杆菌 30S 亚基含有一个 16S rRNA 和 21 种蛋白质（S1～S21）。50S 亚基含有一个 5S rRNA，一个 23S rRNA 以及 34 种蛋白质（L1～L34）。真核生物细胞质核糖体比原核生物的核糖体更大，含有更多的 RNA 和蛋白质。30S 核糖体亚基的体外组装开始于 16S rRNA。然后蛋白质有序地结合上去，这一过程中前面蛋白质的加入帮助后面蛋白质结合于组装颗粒之上。

16S rRNA 的序列研究引发了对这一分子二级结构（分子内碱基配对）的推测。X 射线结晶学研究已经证实这些研究结论。表明 30S 亚基具有一个广泛碱基配对的 16S rRNA，它的形态勾勒出整个颗粒的形态。X 射线结晶学研究也已经证实了大部分 30S 核糖体蛋白的位置。

30S 核糖体亚基具有两种功能：促进密码子和氨酰-tRNA 反密码子之间的正确解码和校正，以及参与蛋白质翻译。结合有 3 种抗生素干扰亚基功能的 30S 核糖体亚基晶体结构使翻译和解码过程清楚地显现出来。壮观霉素结合于 30S 亚基的颈部附近，在那里干扰翻译所必需的头部转动。链霉素结合于 30S 亚基的 A 位点附近，并稳定核糖体的 *ram* 构型。通过使非正确的氨酰-tRNA 相对容易地结合于 A 位点以及通过阻止向校正所必需的严谨构型的转变，降低翻译的忠实度。巴龙霉素结合于 16S rRNA 的 H44 螺旋 A 位点附近的小凹槽内，使碱基 A1492 和 A1493 脱离原位，稳定密码子和反密码子之间的碱基配对，包括非关联氨酰-tRNA 上的反密码子，所以翻译忠实度降低。

结合于 30S 核糖体亚基的 IF1 的 X 射线晶体结构表明 IF1 结合于 A 位点。在这个位置上，它明显地阻断甲酰甲硫氨酸-tRNA 对 A 位点的结合，而且通过一种可能的 IF1 和 IF2 相互作用促进甲酰甲硫氨酸-tRNA 结合于 P 位点。IF1 也和 30S 亚基的螺旋 44 密切相互作用，这可以解释 IF1 怎样促进核糖体亚基的结合和分解。

50S 核糖体亚基的晶体结构已经在 2.4Å 分辨率上被确定。这一结构揭示在核糖体亚基之间的接触面上蛋白质相对很少。用酶过渡态类似物标记的肽基转移酶活性中心 1.8 Å内没有蛋白质。P 位点 tRNA的 2′-OH 处于最佳位置，与 A 位点的氨酰-tRNA形成一个氢键，从而帮助催化肽基转移酶反应。与这一假设一致的是除去该羟基几乎消除了肽基转移酶的全部活性。类似地，除去 23S rRNA 的 A2451 的 2′-OH，可强烈抑制肽基转移酶活性。该羟基可能也通过形成氢键参与催化反应，或通过帮助反应物正确定位参与催化反应。穿过 50S 亚基的通道出口仅够通过一个 α 螺旋蛋白。通道壁由 RNA 构成，其亲水性可能让暴露出疏水侧面的新生蛋白质链容易滑过。

大多数 mRNA 由多个核糖体在同一时间翻译，许多核糖体翻译一条 mRNA 链形成串状的结构就被称为多聚核糖体。在真核生物中，核糖体位于细胞质中。在原核生物中，基因的转录和 mRNA 的翻译同时发生。因此，可发现许多核糖体与某个活性基因结合在一起。

tRNA 作为一种独立于核糖体的小 RNA 被发现，它能负载一个氨基酸，而后将其传给正在合成的多肽。所有 tRNA 都有共同的三叶草形二级结构。它们有 4 个碱基配对茎，即 3 个茎环（D 环、反密码子环和 T 环）和受体茎，用于氨基酸的负载。tRNA也共有一种三维形状，类似于倒置的 L。这种形态通过使 D 茎的碱基同反密码子茎的碱基以及 T 茎的碱基同受体茎的碱基线性排列使分子的稳定性达到最大。tRNA 的反密码子从反密码子环的一侧突出并且被扭曲成一种与 mRNA 中相应的密码子容易配对的形状。

tRNA 的受体茎和反密码子在关联氨酰-tRNA 合成酶的 tRNA 识别过程中起关键作用。在特定情况下，两者中的任何一种都能决定负载的特异性。X 射线晶体结构已经表明合成酶-tRNA 相互作用在两类氨酰-tRNA 合成酶之间不同。class I 合成酶有针对关联 tRNA 的受体茎和反密码子的裂缝，并且从 D 环和受体茎小沟一侧靠近 tRNA。class II 合成酶也有针对受体茎和反密码子的裂缝，但是从反面接近其 tRNA，包括可变臂和受体茎的大沟。

至少有些氨酰-tRNA 合成酶的氨基酸选择性由双筛（double-sieve）机制控制。第一次筛选是粗筛，排除太大的氨基酸，通过合成酶的氨基酸活化位点来完成。这个位点大小足以容纳同类的氨基酸，但不能容纳更大的氨基酸。第二次筛选是细筛，降解过小的氨酰-AMP，通过合成酶与接纳小氨酰-AMP 并水解它们的第二活化位点（编辑位点）完成这一任务。关联氨酰-AMP 由于过大不能嵌入编辑位点，所以它可以避免被水解。

复习题

1. 绘出 *E. coli* 核糖体 30S 和 50S 亚基的草图，显示它们是如何结合形成 70S 核糖体的。

2. 绘出核糖体 30S 和 50S 亚基结合处的草图，指出在 A、P、E 位点中 tRNA 的大概位置。

3. tRNA 的哪一部分与 30S 核糖体亚基结合？哪一部分与 50S 核糖体亚基结合？

4. 在 A 位点和 P 位点反密码子的相互接近为什么很重要？

5. 在 A 位点和 P 位点 tRNA 的受体茎相互接近为什么很重要？

6. 描述本章中介绍的双向凝胶电泳的过程，说明在什么情况下双向凝胶电泳优于单向凝胶电泳。

7. 给出合理的假设，解释以下抗生素如何干扰翻译并给出各假设的证据：a. 链霉素，b. 巴龙霉素。

8. X 射线衍射数据怎样排除了核糖体蛋白在肽基转移酶中作为激活位点？

9. 给出证据说明 P 位点肽酰-tRNA 的末端腺苷的 2′-OH 在转肽中的重要性。该羟基是怎样参与转肽的?

10. 给出证据说明 23S rRNA 的 A2451 的 2′-OH 在转肽中的重要性。该羟基是怎样参与转肽的?

11. 怎样知道碱基 A2451（在 *H. marismortui* 是 A2486）在转肽中不重要?

12. 介绍发现 tRNA 的实验。

13. tRNA 的“三叶草”二级结构是如何发现的?

14. 图示 tRNA 的“三叶草”结构并指出重要的结构元件。

15. 描述实验并给出结果，说明核糖体对氨酰-tRNA 的 tRNA 部分而不是氨基酸部分做出响应。

16. 描述实验并给出结果，说明 tRNA 受体茎的 G3-U70 碱基对是 tRNA 负载丙氨酸的决定因素。

17. 至少给出一条证据，证明氨酰-tRNA 合成酶对 tRNA 识别反密码子的重要性。

18. 根据 X 射线晶体学研究，说明 tRNA 的哪一部分与相应的氨酰-tRNA 合成酶结合?

19. 绘图说明确保氨基酸在氨酰-tRNA 合成酶中选择性的双筛机制。

20. 概括异亮氨酸-tRNA 合成酶双筛选择排除较大和较小氨基酸的证据。

分析题

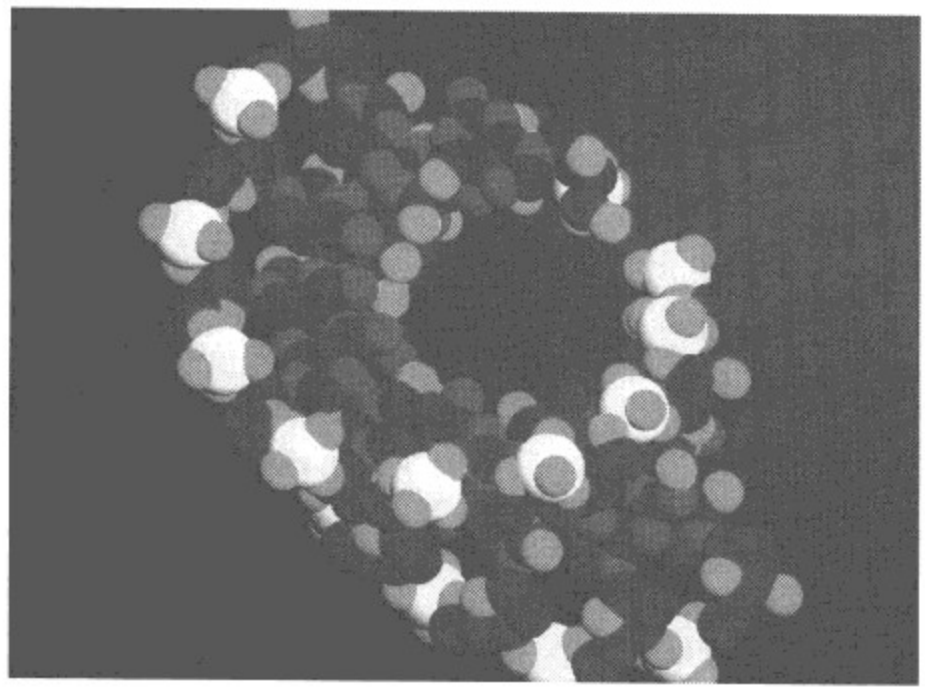
1. 图示真核生物多聚核糖体的初生蛋白链。区分 mRNA 的 5′端和 3′端，并用箭头标出核糖体沿 mRNA 移动的方向。

2. 图示原核生物的基因转录和翻译同时进行。说明初生 mRNA 有核糖体附着，但不显示初生蛋白。用箭头标出转录方向。

3. 你正在研究 $tRNA^{Phe}$，其负载特异性可能受 D 臂的 C11-G24 碱基对影响。设计两个实验证明改变该碱基对会引起 tRNA 负载特异性的改变。第一个实验应该是一个体外生化反应。第二个应该是一个体内遗传反应。

翻译　马　纪　校译　张富春

第 20 章　DNA 复制 I：基本机制与酶学



计算机生成的冈崎片段模型。冈崎片段是 DNA 复制过程中合成的单链短片段。

© *Ken Eward/SS/Photo Researchers, Inc.*

在第 3 章我们已经学习了基因的 3 种主要功能。功能之一是携带遗传信息，我们已经在本书的大部分章节中学习了细胞是如何通过转录和翻译解读遗传信息的。基因的另一个功能是复制，在接下来的两章中将详细分析复制过程。此外，在本章中我们还将学习 DNA 的损伤与修复。

20.1　DNA 复制的一般特征

我们首先了解一下 DNA 复制的一般机制。DNA 的双螺旋模型表明组成 DNA 分子的两条链是互补的，因此，原则上讲，每条链都可以作为模板合成其互补链。正如我们将要学习到的，DNA 复制的半保留模式是正确的。此外，分子生物学家还揭示了 DNA 复制的另一些有趣的一般性特征：半不连续复制（形成短片段然后连接起来）、需要 RNA 引物、双方向复制。下面，我们依次来了解这些特征。

半保留复制

DNA 复制的沃森-克里克模型（已在第 2 章介绍）认为，DNA 分子中新合成的链遵循常规的 A-T、G-C 碱基配对原则。该模型还指出：DNA 分子中的两条亲本链之间的氢键断裂而彼此分开，各自作为模板链合成一条与之互补的新生子链，新生的互补链与亲本链构成子代 DNA 分子。这种复制方式称为半保留复制（semiconservative replication），因为每个子代双螺旋 DNA 分子中，一条链是亲本链，另一条链是子链（图 20.1a）。也就是说，在每一个子代 DNA 分子中都保留有一条亲本链。然而，这并不是 DNA 复制唯一的可能方式，另一种可能的机制是全保留复制（conservative replication）（图 20.1b），两条亲本链彼此结合，恢复原状，新合成的两条子链彼此互补结合形成一条新的双链。当然，还可能采取随机弥散型复制（random dispersive replication）方式，亲代双链被切成双链片段，复制完成后，“新旧”DNA 同时存在于同一条链中（见图 20.1c），这种复制机制被认为可以避免双链 DNA 解旋的问题。

1958 年，Matthew Meselson 和 Franklin Stahl 通过一个经典的实验对 DNA3 种复制的可能性进行了辨别。研究者通过在添加氮同位素的培养基中培养 *E. coli* 细胞的方法而使 *E. coli* 的 DNA 被 ^{15}N 所标记，^{15}N-DNA 的密度比普通 ^{14}N-DNA 的密度大，然后，将在含有 ^{15}N 的培养基上生长的 *E. coli* 转接到含 ^{14}N 的普通培养基中培养不同时间，最后，通过 CsCl 密度梯度超速离心来判定 DNA 的密度。图 20.2 描述了对照实验的结果，表明用这种方法可以将 ^{15}N-DNA 和 ^{14}N-DNA 清晰地区分开来。

如果按照这 3 种不同的复制机制，DNA 复制一轮后会出现怎样的结果呢？如果是全保留复制，则

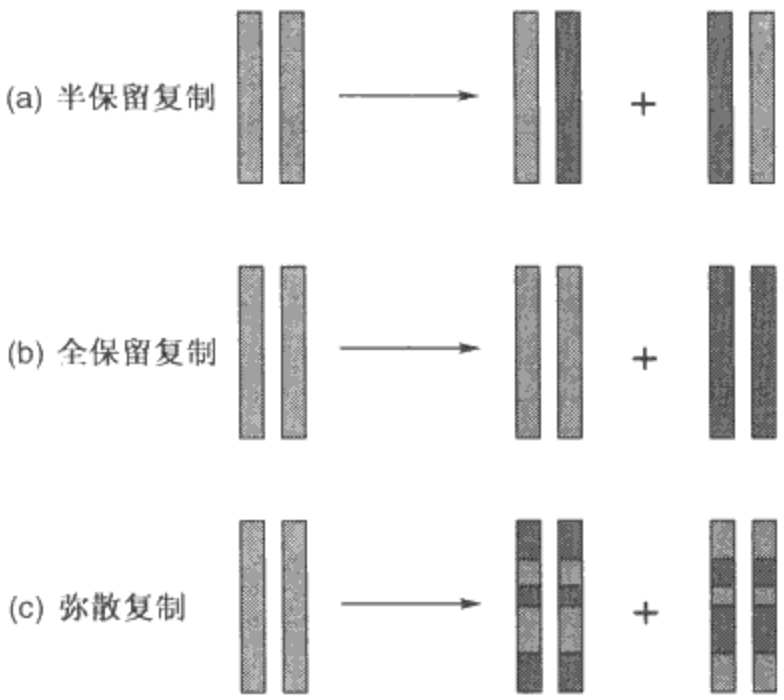


图 20.1　DNA 复制的 3 种假说。(a) 半保留复制产生两个子代双螺旋 DNA 分子，每个子代 DNA 分子中都保留有一条旧链（蓝色）和一条新链（红色）；(b) 全保留复制产生两个子代双螺旋 DNA 分子，其中一个 DNA 分子由两条旧链组成（蓝色），另一个 DNA 分子由两条新合成的子链组成（红色）；(c) 弥散型复制产生两个子代双螺旋 DNA 分子，每一个子代双螺旋 DNA 分子中的同一条链是由旧链和新链混合形成的。

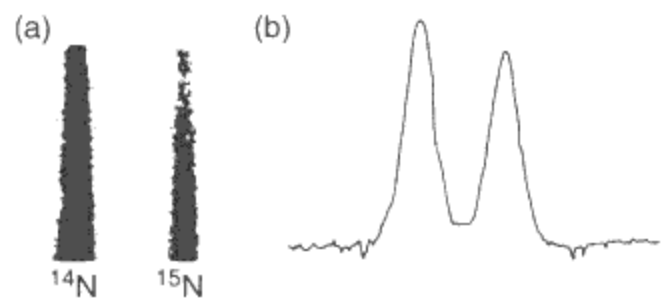


图 20.2　通过 CsCl 密度梯度超速离心方法分离 DNA。将含普通氮同位素（^{14}N）的 DNA 与重氮同位素（^{15}N）标记的 DNA 混合后进行 CsCl 密度梯度超速离心，由于 ^{15}N-DNA 和 ^{14}N-DNA 的密度不同，通过 CsCl 密度梯度离心而分成两条清晰的条带。(a) 离心管在紫外光照射下的图像。此图像是在离心机旋转的情况下通过其窗口观察到的，该离心机的设计可以让实验者在不停止离心机旋转的情况下观察离心的样品。两黑色条带分别代表了能够吸收紫外线的两种不同 DNA。(b) 每一条带黑度的曲线图。表示了两种 DNA 的相对含量。(*Source*: Adapted from Meselson, M. and F. Stahl, The replication of DNA in *Escherichia coli*. *Proceedings of the National Academy of Sciences USA* 44 (1958) p. 673, f. 2.)

两条重的亲本链仍然结合在一起，那么就会产生一个新的双螺旋 DNA 分子。因为这个新的双螺旋 DNA 分子含有轻的 ^{14}N，两条链都具有较小的密度。这样由两条重链（H/H）组成的亲本双螺旋 DNA 分子与由两条轻链（L/L）组成的子代双螺旋 DNA 分子就很容易通过 CsCl 密度梯度离心而分开（图 20.3a）。另一方面，如果是半保留复制，则亲代 DNA 分子中的两条重链就会彼此分开，用于合成新的轻的与之互补的子链。这种由一条重链和一条轻链（H/L）组成的杂交双螺旋 DNA 分子的密度将介于完全由重链（H/H）组成的双链 DNA 分子的密度与完全由轻链（L/L）组成的双链 DNA 分子的密度的中间值（图 20.3b）。图 20.4 表明复制事件确实是以半保留方式进行的：在第一轮 DNA 复制完成后，经 CsCl 密度梯度离心，在离心管中只出现一条带，位于由重链 DNA 形成的条带和由轻链 DNA 形成的条带的中间位置。这就排除了全保留复制的可能性，但弥散型复制也会出现这样的结果。

当 DNA 再复制一轮后，其结果就会排除弥散型复制假设的可能性。当 *E. coli* 细胞在含有 ^{14}N 的培养基上复制两轮后，按照弥散型复制机制，DNA 复制产物中重链占 25%、轻链占 75%。半保留复制则产生一半 H/L 类型的 DNA 分子和一半 L/L 类型的 DNA 分子（见图 20.3b）。换句话说，首轮复制产生的 H/L 杂交分子，其两条链分开，分别作为模板用于合成互补的子链，产生的 H/L 类型 DNA 分子与 L/L 类型 DNA 分子的比例为 1∶1。并且事实确实如此（图 20.4）。为了证实这一中间密度峰值确实是由重链和轻链按 1∶1 组成的 DNA 分子产生的，Meselson 和 Stahl 将完全由 ^{15}N 标记的 DNA 分子与在 ^{14}N 培养基上繁殖 1.9 代的 DNA 混合在一起，然后测定峰值间的距离。中间的峰值几乎完全地处于另外两峰值的中心位置（两峰值间距的 50% ± 2%）。因此，这一研究数据是对 DNA 半保留复制机制的强有力支持。

小结　DNA 分子以半保留方式进行复制，当亲本链分离后，分别作为模板合成一条新的与之互补的子链。

半不连续复制

如果让我们设计 DNA 的复制机制，可能我们想到如图 20.5a 所示：DNA 双链解旋形成复制叉，沿着复制叉移动方向，两条新 DNA 子链以相同方向连续合成。然而，这一方案具有致命的缺陷：它要求这一复制系统能够按 5′→3′及 3′→5′ 两个方向来合成 DNA，因为在 DNA 分子中两条链是反向平行的，如果一条按 5′→3′ 方向由左向右，那么另一

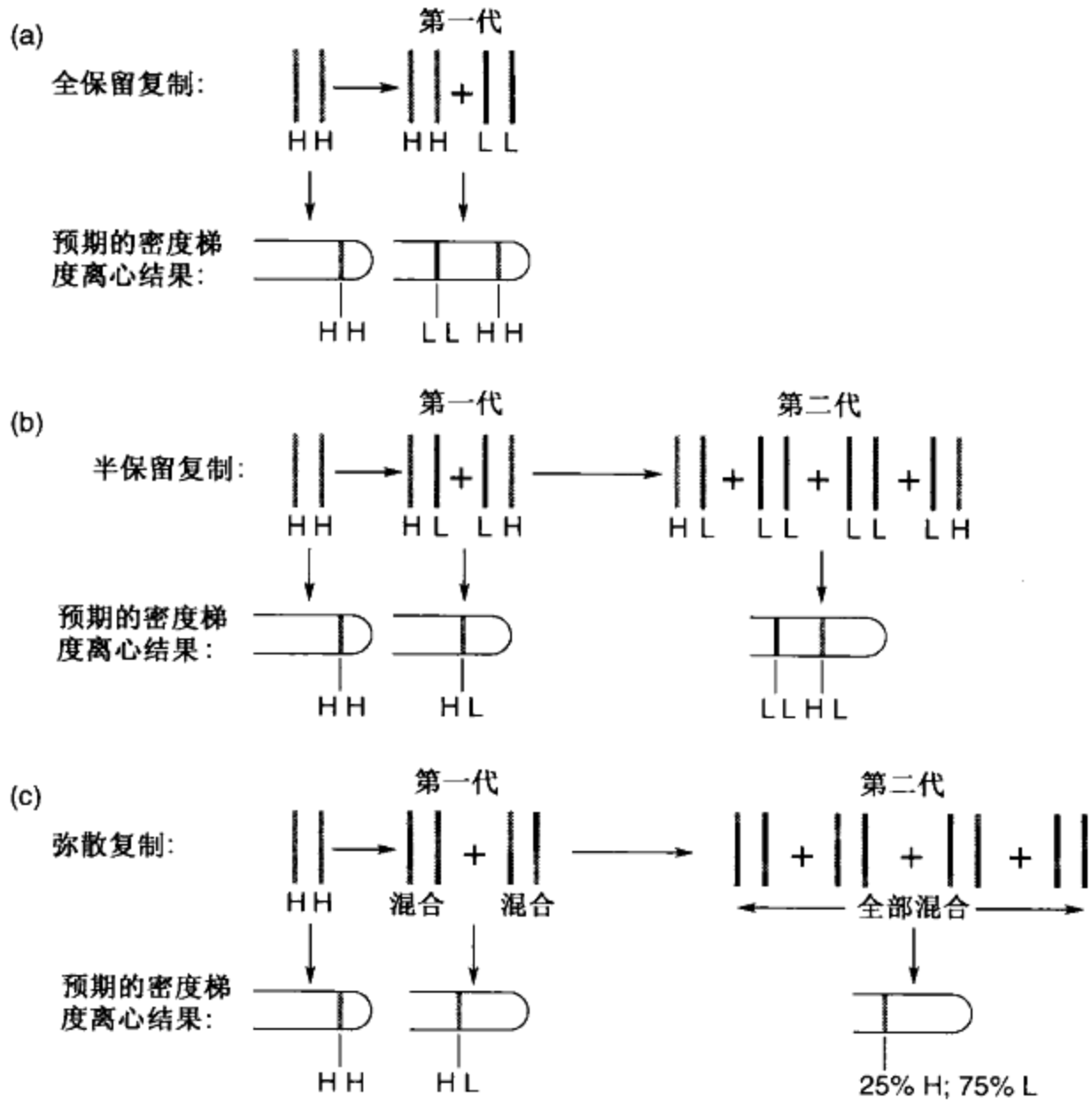


图 20.3　DNA 复制的 3 种假设。在细菌繁殖一代即 DNA 复制一轮后，全保留复制模型（a）预期会产生数量相同的两种不同 DNA 分子，即一种是完全由重链组成的双链 DNA 分子（H/H），另一种是完全由轻链组成的双链 DNA 分子（L/L）。半保留复制模型（b）和弥散型复制模型（c）均预期会出现一条带，其 DNA 分子密度介于重链 DNA 分子和轻链 DNA 分子的中间。Meselson 和 Stahl 的实验结果证实了后一种预测，因此排除了全保留复制的可能性。按弥散型复制，当细菌繁殖两代后，仍然产生单密度条带，在所有 DNA 分子中，重链占 25%，轻链占 75%。其 DNA 分子的密度介于轻链 DNA 分子和由一条轻链与一条重链组成的 DNA 分子中间。半保留复制则会产生数量相同的两种不同的 DNA 分子，即一种是完全由轻链组成的双链 DNA 分子（L/L），另一种是由一条轻链与一条重链组成的双链 DNA 分子（H/L）。实验结果再次证实了后一种推测，即 DNA 半保留复制是完全正确的。

条链一定是按 3′→5′由左向右。但是所有天然的 DNA 复制系统的合成部件（DNA 聚合酶）只能按一个方向即 5′→3′方向合成 DNA 子链。也就是说，在 5′端插入第一个核苷酸后，通过在生长链的 3′端不断添加核苷酸而使子链不断向 3′端延伸。

按照这一推理，Reiji Okazaki 推断 DNA 的两条链并不完全都是连续复制的，从理论上讲，DNA 聚合酶能够按 5′→3′ 方向连续合成一条子链（前导链，leading strand），而另一条子链（后随链，lagging strand）的复制一定是不连续的，如图 20.5b 和图 20.5c 所示。后随链之所以发生不连续复制，是因为其合成方向与复制叉移动方向相反，当复制叉展开暴露出新的区域用于 DNA 复制时，后随链就沿着远离复制叉的“错误”方向生长，新暴露出的 DNA 区域只能在复制叉处以重新起始的方式进行复制，而复制叉后面的 DNA 片段已完成复制。这种 DNA 合成的起始与再起始反复不断地发生。当然，新产生的短 DNA 片段一定会以某种方式连接起来形成一条连续的 DNA 链，成为 DNA 复制的最终产物。

Okazaki 的半不连续复制模型提出的两个预测均得到其研究团队的实验验证：①因为至少有一半新合成的 DNA 首先是以短的片段形式出现的，用放射性 DNA 前体经脉冲标记，在连接前就能够标记

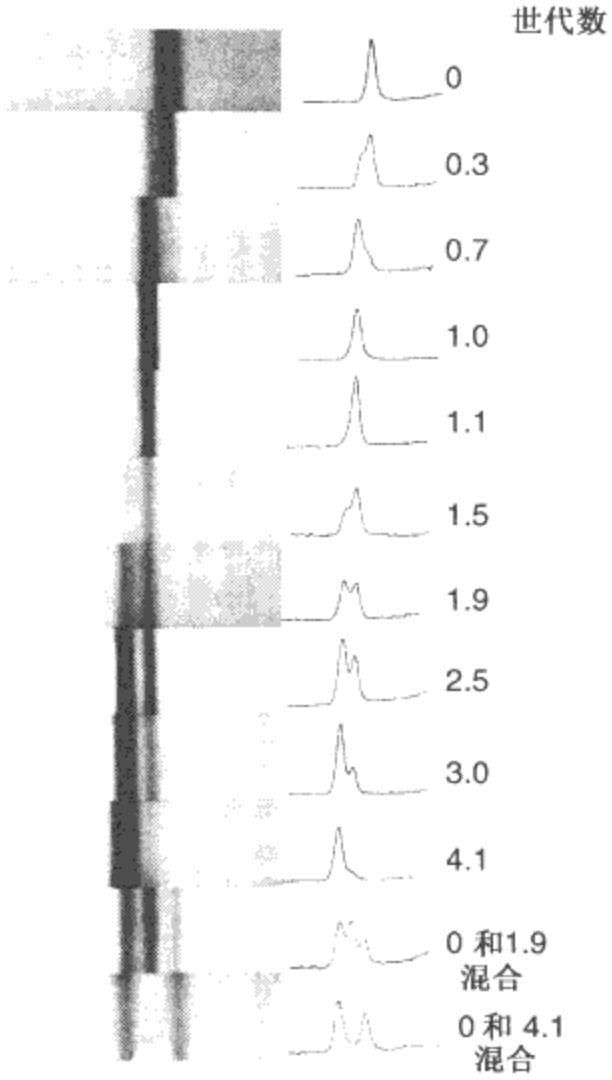


图 20.4　证明 DNA 复制是半保留复制的 CsCl 密度梯度超速离心实验结果。Meselson 和 Stahl 将用^{15}N 标记的 *E. coli* 细胞转接到含有^{14}N 的培养基上，并控制其繁殖的代数，繁殖代数在图版右侧标出，然后将细菌 DNA 进行 CsCl 密度梯度超速离心。(a) 旋转离心管在紫外照射下形成的图像，暗带对应于重链 DNA（右）和轻链 DNA（左）。在这两条带的中间也能观察到中间密度值的条带，在 1.0 代和 1.1 代时确实只观察到这一条中间带，这条带对应的双螺旋 DNA 分子是由一条被^{15}N 标记的重链和一条^{14}N 轻链组成，正如半保留复制模型所预测的那样。当繁殖到 1.9 代时，Meselson 和 Stahl 观察到峰值大致等量的两条带，即中间带 H/L 及 L/L 带，这也是半保留复制模型所预测的结果。当繁殖到 3 代或 4 代以后，他们观察到，中间带 H/L 逐渐消失，而 L/L 带却逐渐增加，这也是半保留复制模型所预测的结果。(b) 利用密度计分析图 (a) 中的条带，量化每一条带 DNA 的量。(*Source*: Meselson, M. and F. W. Stahl, The replication of DNA in *Escherichia coli*, *Proceedings of the National Academy of Sciences USA* 44: 675, 1958.)

并捕捉到这些短片段；②如果消除具有连接功能的 DNA 连接酶 (DNA ligase)，那么用放射性 DNA 前体进行相对较长时间的脉冲标记，也能检测到这些短的 DNA 片段。

Okazaki 选择 T4 噬菌体的 DNA 复制作为其研究 DNA 复制的模式系统，该系统具有简单且易获得 DNA 连接酶突变体的优势。为验证第一个推测，Okazaki 及同事在逐渐缩短脉冲标记时间的情况下，用^{3}H-脱氧胸苷脉冲标记进行 T4 DNA 复制的 *E. coli* 细胞。为确保这些短片段在连接前被捕捉，研究者采用 2s 的脉冲标记时间。最后，通过超速离心方法测定了新合成 DNA 片段的大小。

图 20.6a 显示了实验的结果。脉冲标记 2s 后，通过密度梯度离心，能够观测到一些标记 DNA，在极限检测范围内，所有标记 DNA 都是很小的片段，长度为 1000～2000nt，这些片段位于离心管顶部附近。随着脉冲时间的延长，在靠近离心管底部位置处出现了另一个标记 DNA 的峰，这是由于新合成的标记 DNA 小片段连接到在标记实验开始前就已合成的 DNA 大片段上而造成的结果。因为在实验开始前这些大片段并没有被标记，所以不会显现出来，只有经过一段足以使连接酶将较小的标记片段连接到这些大片段上的时间，才能显现，这只需几秒钟的时间。他们恰如其分地将复制时最初形成的 DNA 短片段称为冈崎片段 (Okazaki fragment)。

冈崎片段的发现为 T4 DNA 在复制过程中至少是部分采取不连续复制的方式提供了证据。当 DNA 连接酶失去功能时，小的 DNA 片段就会积聚到一个很高的水平，这充分证明这一推测。Okazaki 研究小组用 DNA 连接酶基因缺陷型 T4 突变体开展了该实验的研究工作。图 20.6b 显示，在该突变体中冈崎片段的峰值占主导。即使进行 1min 的长时间标记，依然存在大量的标记片段，这表明冈崎片段并不是短时脉冲标记的产物。

只有 DNA 的两条链均以不连续方式进行复制才可以解释小片段标记 DNA 显著积累的现象，如图 20.5c 所示。但这些小片段 DNA 产生的真正原因可能是 DNA 修复系统将错配的 dUMP 残基切除的结果。UTP 是 RNA 的一个基本前体，但细胞内也存在 dUTP，会偶尔取代 dTTP 错配掺入到 DNA 链中（以 dUMP 的形式）。细胞内有两种酶能够减少这种情况的发生：一种是由 *dut* 基因编码的 dUTP 酶 (dUTPase)，能够降解 dUTP；另一种是由 *ung* 基因编码的尿嘧啶 *N*-糖苷酶 (uracil *N*-glycosylase)，能够从 DNA 链中将错误掺入的尿嘧啶切除而产生一个无碱基位点，在修复过程中该位点易被水解而断裂。这样，不管是连续复制还是不连续复制，都会因修复而产生一些短片段。问题是，在图 20.6 中观测到的冈崎片段中，不连续复制所产生的冈崎片段与因修复错配的 dUMP 残基所产生的片段

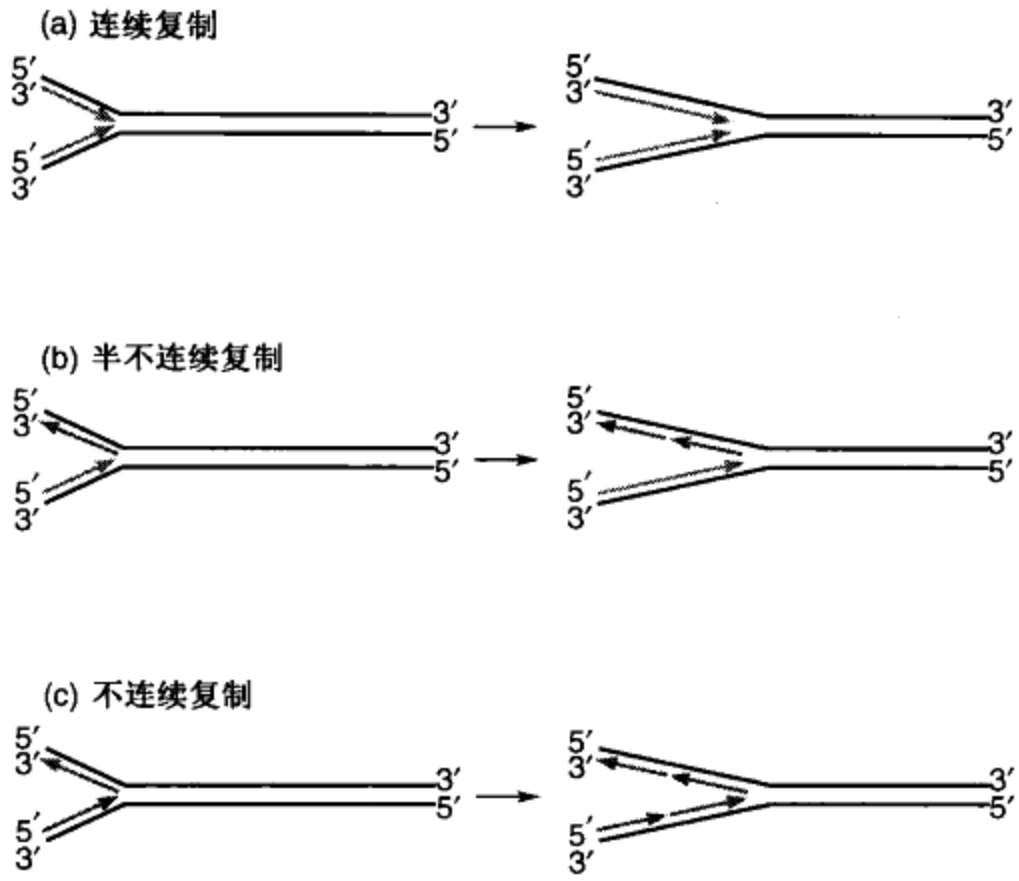


图 20.5　DNA 复制的连续、半不连续和不连续模型。（a）连续复制模型。当复制叉向右移动时，两条链由左向右沿同一方向进行连续复制（蓝色箭头），上方链按 3′→5′方向合成，下方链按 5′→3′方向合成；（b）半不连续复制模型。其中的一条新链（前导链，下方）是连续合成的（蓝色箭头），同（a）图所示。而另一条新链（后随链，上方）是不连续合成的（红色箭头）；两条链均按 5′→3′方向合成；（c）不连续复制模型。前导链和后随链均合成短片段（不连续，红色箭头），两条链均按 5′→3′方向合成。在生物体内，DNA 复制是半不连续的，如图（b）所示的半不连续复制模型。

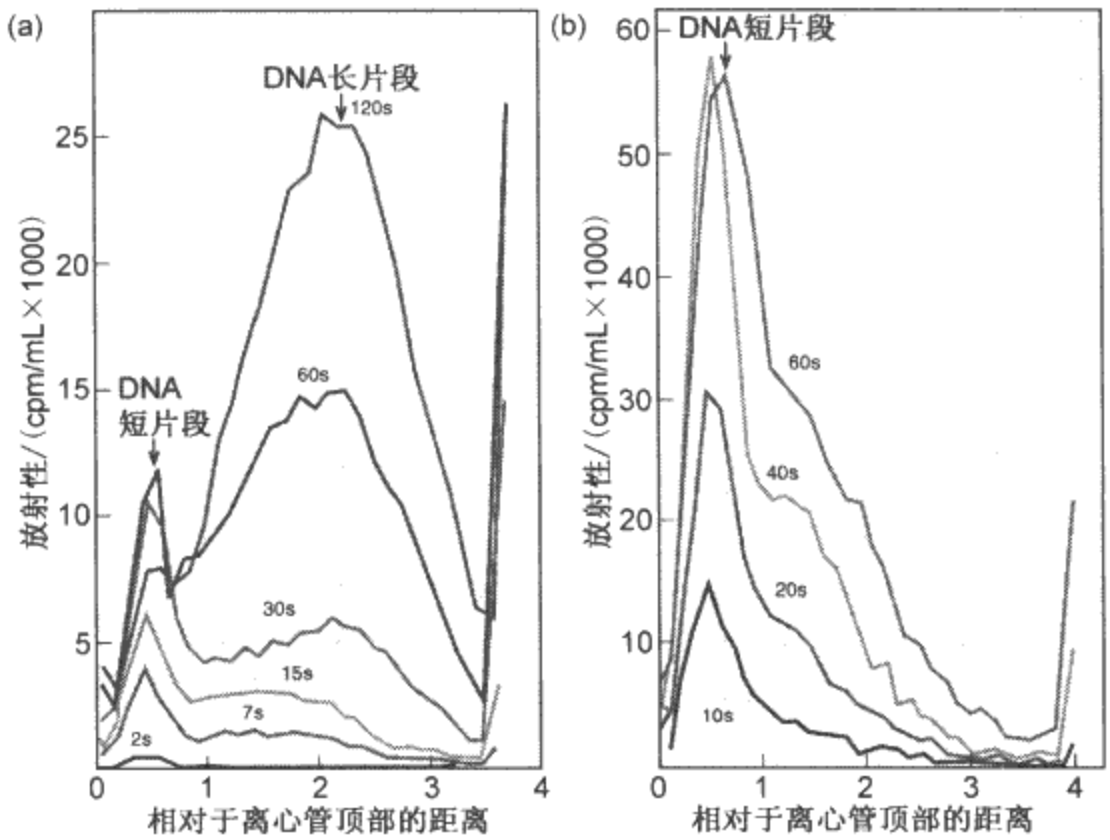


图 20.6　DNA 复制至少是半不连续的实验证据。（a）冈崎及同事用放射性前体以极短的脉冲时间标记复制的 T4 DNA，通过密度梯度超速离心，将 DNA 产物按大小进行分离，在脉冲标记时间最短时，主要是短片段获得了标记（位于离心管顶部附近），与不连续复制模型的预期结果相符；（b）在DNA 连接酶基因缺陷的突变体中，即使在较长的脉冲标记时间（结果中显示的是 1min）条件下，仍然有 DNA 短片段的累积。

各占多少比例？随后的研究结果表明：因修复而产生的短片段约占冈崎片段总量的一半，所以 DNA 复制是以半不连续方式进行的：一条链连续复制，而另一条链不连续复制。

小结　在 *E. coli*（以及其他生物体）中 DNA 复制是半不连续的，一条链是沿复制叉移动的方向连续复制，而另一条链是沿相反方向进行不连续复制，产生 1～2kb 的冈崎片段，从而使两条链都按 5′→3′方向合成 DNA 子链。

DNA 合成的引发

在前面章节中我们已经了解到，RNA 聚合酶能够很简单地通过从头开始一条新生 RNA 链的合成而起始转录，RNA 聚合酶将第一个核苷酸引入正确的位置，然后将第二个核苷酸添加在第一个核苷酸的 3′端，但 DNA 聚合酶却不能以同样的策略起始 DNA 的合成。如果我们在一个反应体系中加入 DNA 聚合酶、所有的核苷酸以及合成 DNA 所需要的其他小分子，并加入一条单链 DNA 或没有断裂的双链 DNA 后，DNA 聚合酶并不能合成一条新的 DNA 链，这是为什么？难道缺少某种成分？

现在已经知道缺少的成分是引物，是能够被 DNA 聚合酶捕获的核苷酸片段，并能向其 3′端添加核苷酸而使引物延长，这段引物不是 DNA，而是

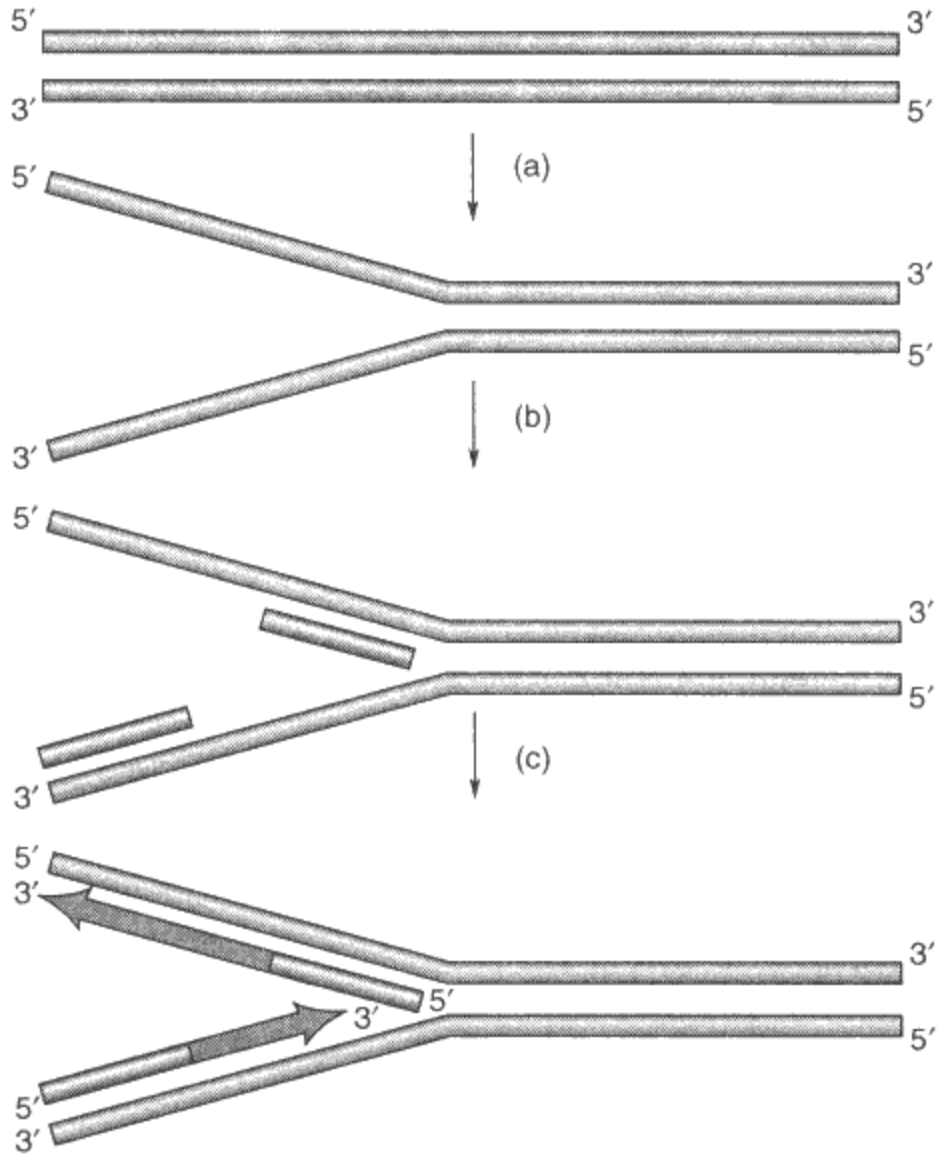


图 20.7　DNA 合成的引发。（a）两条母链（蓝色）分离；（b）合成短的 RNA 引物（红色）；（c）DNA 聚合酶以引物为起始点合成子代 DNA 链（绿色箭头）。

RNA 小片段。图 20.7 为该过程的简化图示。首先，复制叉打开；其次，短 RNA 引物的合成；再次，DNA 聚合酶通过在引物上添加脱氧核糖核苷酸而合成 DNA，如图 20.7 中箭头所示。

支持 RNA 引发 DNA 合成的第一个证据是抗生素利福平能够抑制*E. coli* 细胞提取物对 M13 噬菌体 DNA 的复制。令人奇怪的是，利福平能够抑制 *E. coli* RNA 聚合酶的活性，但不会抑制 DNA 聚合酶的活性，对此情况的解释就是 M13 噬菌体利用 *E. coli* 的 RNA 聚合酶合成用于 DNA 复制的 RNA 引物，然而，这并不是普遍现象。即使是*E. coli* 也不会利用自身的 RNA 聚合酶来引发 DNA 的合成，而是通过专门的酶系实现 DNA 合成的引发。

RNA 引发 DNA 复制的最有力证据是 DNA 酶不能完全降解冈崎片段，总是剩余一些长度为10～12bp 的 RNA 小片段。这项研究工作主要是由 Reiji Okazaki 的夫人 Tuneko Okazaki 及同事完成的，Tuneko Okazaki 及同事最初估算的 RNA 引物太小，长度仅为 1～3nt。有 2 个难题造成了引物的长度被低估：①在提纯之前，RNA 引物已被核酸酶降解；②研究者无法区分降解引物与完整引物。在 1985 完成的第二组实验中，冈崎研究组解决了这两个难题，并发现完整引物的实际长度为10～12nt。

为降低核酸酶活性，研究者以核酸酶 H 缺失的突变菌株、DNA 聚合酶 I 的核酸酶活性缺失的突变菌株和核酸酶 H 与 DNA 聚合酶 I 的核酸酶活性均缺失的突变菌株为对象开展研究，结果发现在这些突变菌株中完整 RNA 引物的产量大幅提高。为只标记未降解的完整引物，研究者用加帽酶、鸟苷酰转移酶及［α-^{32}p］GTP 标记 RNA 引物的 5′端，在第 15 章已介绍过鸟苷酰转移酶能够将 GMP 添加到 RNA 的 5′端磷酸盐上（理论上是末端二磷酸）。如果引物在 5′端被降解，那么它就会失去这些磷酸盐，从而也就失去了标记。

用这种方法将 RNA 引物标记后，研究者用 DNA 酶处理冈崎片段，去除其 DNA 部分，然后，将剩余的标记引物进行凝胶电泳。图 20.8 描述了实验的结果，所有突变菌株中的 RNA 引物均产生了清晰可见的条带，长度约为（11±1）nt。野生菌株没有产生可检测的条带，核酸酶显然已经将大多数或全部完整的 RNA 引物进行了降解，进一步的实验将图 20.8 中的宽带分解为 3 条大小分别为 10nt、11nt 和 12nt 的离散条带。

小结　*E. coli*细胞内的冈崎片段的合成由长度为 10～12nt 的 RNA 引物引发。在野生型细胞内，由于降解 RNA 的酶的存在，很难检测到完整的 RNA 引物。

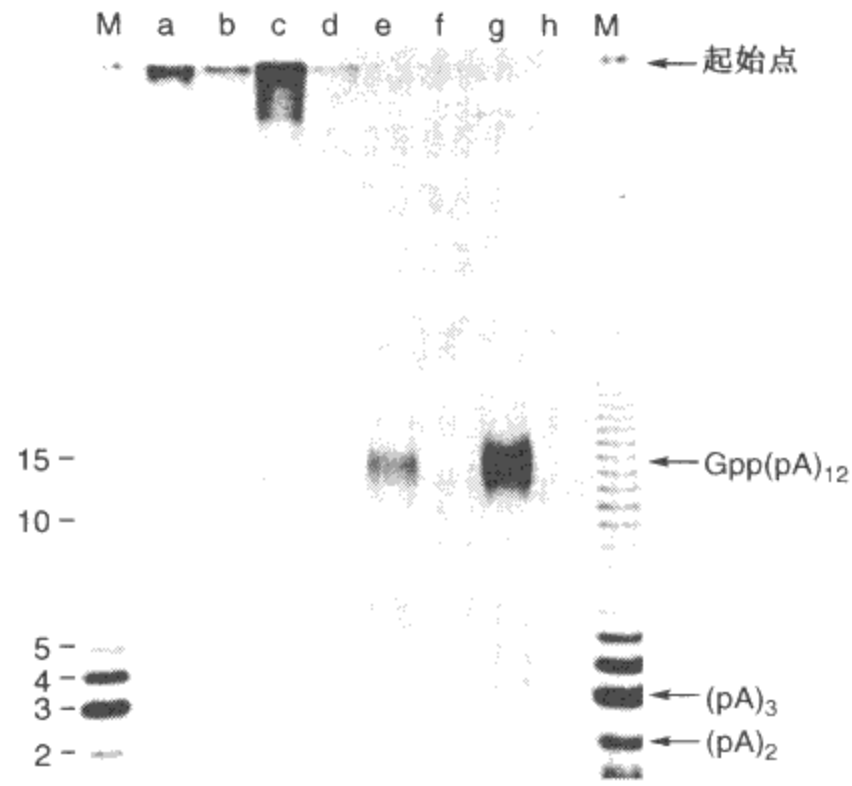


图 20.8　RNA 引物的发现与测定。Tuneko Okazaki 及同事从*E. coli*的野生型菌株和突变型菌株中分离冈崎片段，突变菌株分别缺失降解 RNA 引物的核酸酶中的一种或两种。用［α-^{32}P］GTP 和加帽酶标记冈崎片段中的完整引物，然后，研究者用 DNA 酶降解冈崎片段中的 DNA，对留下的标记引物进行电泳和放射自显影分析。M 泳道为 Marker，a～d 泳道为 DNA 酶降解前，e～h 泳道为 DNA 酶降解后，a 和 e 泳道为 RNase H 缺陷菌株，b 和 f 泳道为 DNA 聚合酶 I 的核酸酶缺陷菌株，c 和 g 泳道为 RNase H 和 DNA 聚合酶 I 核酸酶均缺陷菌株，d 和 h 泳道为野生型菌株。两种核酸酶均缺陷的菌株能产生较多引物（g 泳道），所有处理下引物的长度是(11±1) nt，13 聚体Gpp（pA）$_{12}$ 的位置标注在图的右侧。（*Source*：Kitani，T.，K.-Y. Yoda，T. Ogawa and T. Okazaki，Evidence that discontinuous DNA replication in *Escherichia coli* is primed by approximately 10 to 12 residues of RNA starting with a purine. *Journal of Molecular Biology* 184（1985）p. 49，f. 2，by permission of Academic Press.）

双向复制

20 世纪 60 年代初，John Cairns 用放射性 DNA 前体标记正在复制过程中的*E. coli* DNA 分子，图 20.9a 显示了实验的结果及 Cairns 对实验结果的阐释。图 20.9a 中描述的结构与希腊字母 θ 在外形上类似因而被称为 θ 结构，因为图 20.9a 中 DNA 的 θ 结构并不十分明显，所以在图 20.9b 中简略图示了第二轮复制过程的放射自显影结果。图 20.9 显示，DNA 的复制起始于一个“泡”的产生，即亲本链分离、子代 DNA 开始合成的小区域。随着泡的扩展，正在复制的 DNA 开始呈现 θ 轮廓。现在我们可以辨认图 20.9b 中部示意图所显示的放射自显影图像，此时，θ 的横臂已生长到足够的长度，从而延伸到环形 DNA 分子区域。

图 20.9 所示的 θ 结构中包含有 X 和 Y 两个复制叉（replicating fork），由此产生一个重要的问题：究竟是其中之一还是这两个复制叉均代表 DNA 活跃复制的位点？换句话说，DNA 是单向复制的还是双向复制的？单向复制时，一个复制叉向远离另一个复制叉的方向移动；双向复制时，两个复制叉反向远离原点的方向移动。Cairns 的放射自显影图像无法对这个问题做出回答，但 Elizabeth Gyurasits 和 R. B. Wake 在研究枯草杆菌（*Bacillus subtilis*）DNA 复制时明确发现，DNA 的复制是双向进行的。

研究者采取的研究策略是先使枯草杆菌细胞在含有弱放射性 DNA 前体的培养基上进行短时培养，然后再转接到含强放射性 DNA 前体的培养基上进行短时培养，标记前体均为［^{3}H］胸腺嘧啶。在这一类型的放射自显影分析实验中，氚（^{3}H）是十分有用的放射性同位素，因为其放射性较弱，以至于产生的射线在感光乳剂中停止并产生银色痕迹前不会向远处传播，这意味着放射自显影的银色图像与放射性 DNA 的形状密切相关，更为重要的是未标记的 DNA 并没有出现在放射自显影图像中。实验中的脉冲标记时间很短，以至于只有复制泡才是可见的（图 20.10a），注意不要与图 20.9 中的完整细菌染色体相混淆。

如果仔细观察图 20.10a，你会注意到银色痕迹是不均一的，这些银色痕迹集中在复制泡的两个分叉处，额外的标记表明这一区域的 DNA 是在高放射强度脉冲标记期间复制的，由于都有额外标记，说明两个复制叉在高放射强度脉冲标记期间均处于活动状态。所以，枯草杆菌 DNA 的复制是双向进行的，产生于固定复制原点（origin of replication）的两个复制叉沿环形 DNA 分子相向移动，直至在环形分子的另一侧相遇。使用该方法及其他技术，在后继实验中证明*E. coli* 染色体也是双向复制的。

J. Huberman 和 A. Tsai have 以真核生物果蝇（*Drosophila melanogaster*）为研究对象，开展了同样的放射自显影研究工作。研究者用强放射性（高比度放射性）DNA 前体进行脉冲标记，随后用低放射性（低比度放射性）DNA 前体进行脉冲标记。这两种标记交替进行，即将标记过程颠倒，先进行低

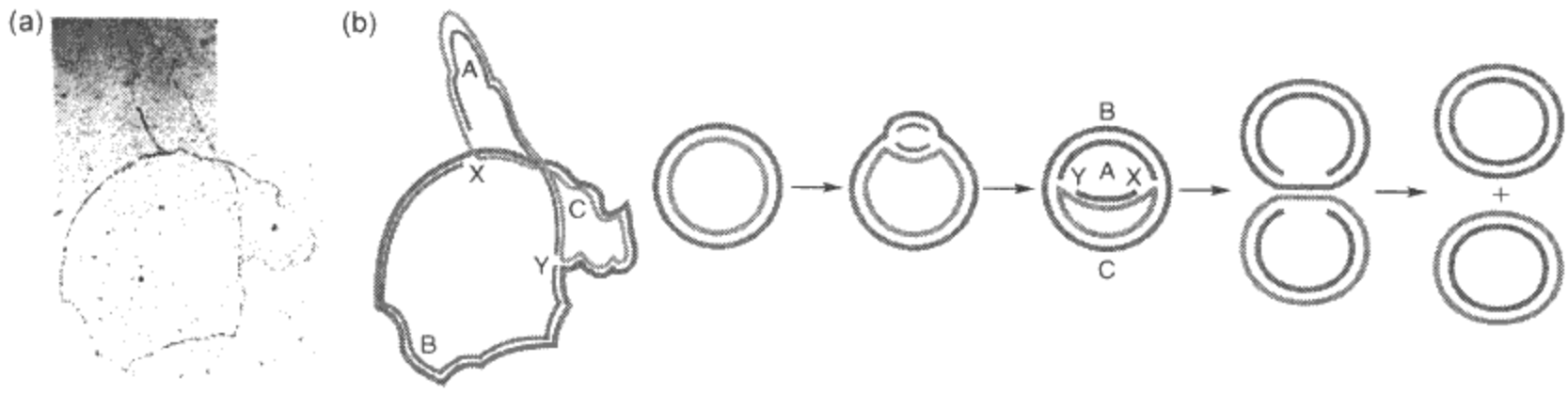


图 20.9　*E. coli* DNA 复制的 θ 模型。（a）正在复制过程中的*E. coli* DNA 放射自显影图像及其图例解释。在放射性核苷酸存在的条件下，使 DNA 完成第一轮复制并进入到第二轮复制。在右侧的解释简图中，红色线代表标记的 DNA，蓝色线代表未标记的亲本 DNA。（b）DNA 复制 θ 模型的详细描述。有色线条的代表含义与图（a）相同。（*Source*：（a）Cairns，J.，The chromosome of *Escherichia coli*. *Cold Spring Harbor Symposia on Quantitive Biology* 28（1963）p. 44.）

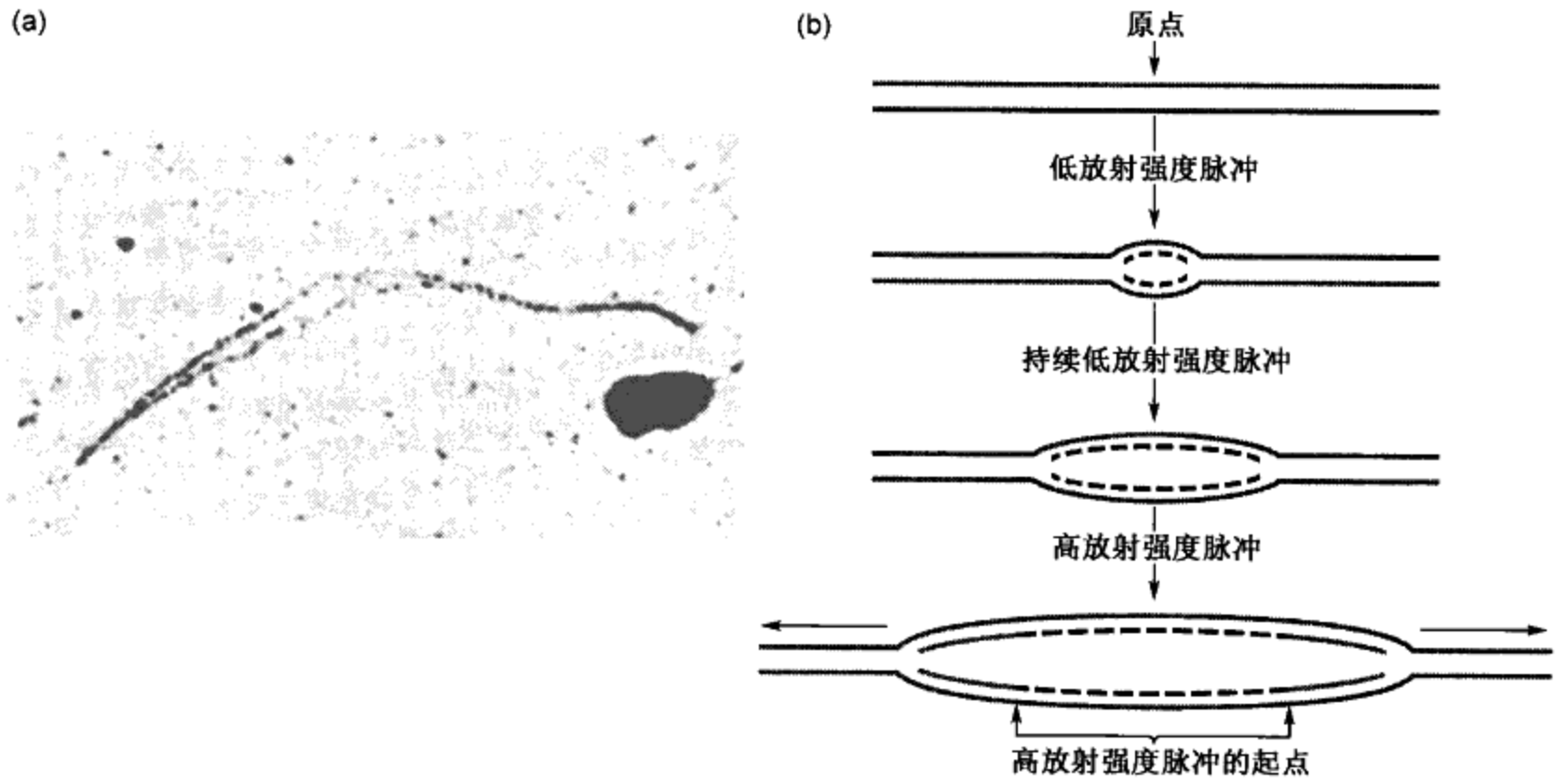


图 20.10　DNA 双向复制的实验论证。（a）正在复制的枯草杆菌（*Bacillus subtilis*）DNA 的放射自显影图像。休眠的细菌孢子在含弱放射性 DNA 前体的培养基上萌发，新形成的复制泡被微弱标记，当复制泡生长到一定长度后，添加放射性较强的 DNA 前体进行短时标记；（b）放射自显影实验结果的阐释。紫色线条代表在低放射性脉冲标记期间产生的弱标记 DNA 链，橙色线条代表在随后的强放射性脉冲标记期间产生的强标记 DNA 链。由于两个复制叉都有强放射性标记，说明枯草杆菌 DNA 的复制是双向的。（*Source*：（a）Gyurasits，E. B. and R. J. Wake，Bidirectional chromosome replication in *Bacillus subtilis*. *Journal of Molecular Biology* 73（1973）p. 58，by permission of Academic Press.）

比度放射性标记，然后在进行高比度放射性标记。将标记的果蝇 DNA 进行放射自显影处理，获得的图像显示，延伸的 DNA 并没有使复制泡保持开放状态，而是以单条银色痕迹的形式出现，且忽而消失、忽而显现。

条斑的一端是标记的开始，另一端则意味着标记的结束。在实验中，标记条斑是成对出现的（图 20.11a），代表着产生于共同起始位点且反向移动的复制叉。为什么与枯草杆菌 DNA 的实验不同，标记并没有从复制原点的中间位置处开始呢？在对枯草杆菌开展的实验中，研究者通过孢子在同一时间萌发而使细胞达到同步化状态，因此，研究者可以在细胞起始 DNA 复制前（如孢子萌发前）将标记加入到细胞内。但在果蝇的实验中，很难能够使细胞达到同步化状态。结果发现，DNA 的复制通常在标记物加入前就已开始了，所以中间的空白区域表

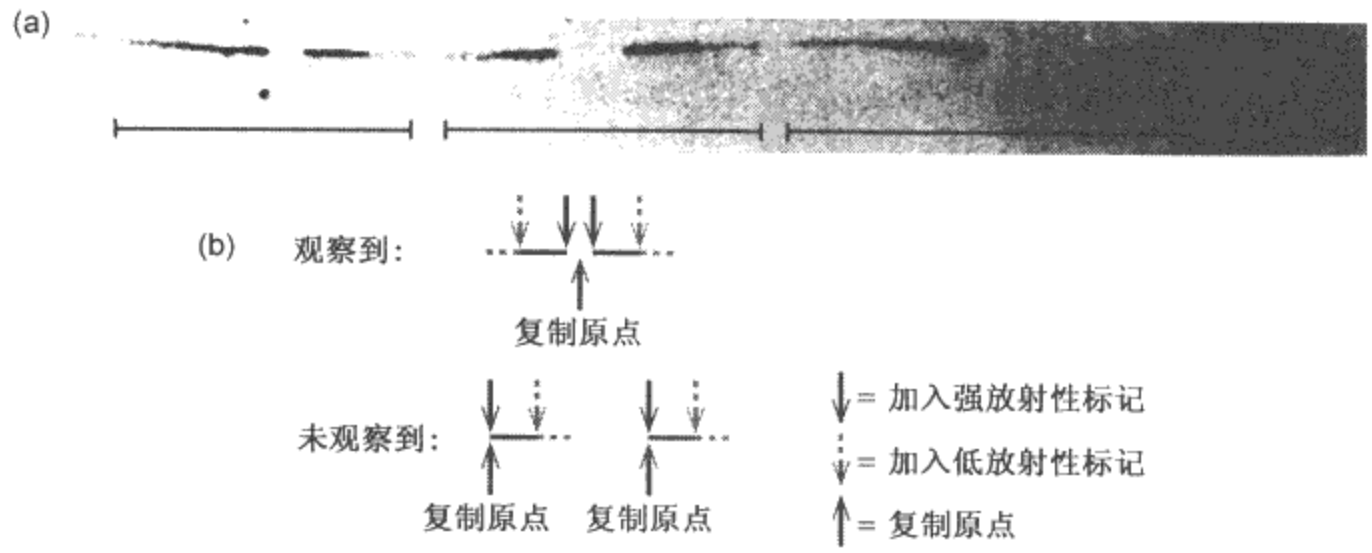


图 20.11　真核生物细胞内的 DNA 双向复制。(a) 正在复制中的果蝇 DNA 放射自显影图像，先用强放射性 DNA 前体进行脉冲标记，随后用弱放射性 DNA 前体进行脉冲标记。注意成对条斑（用方括号表示）从中间向外侧端逐渐变细，这表明被标记的复制子具有一个中央复制起点和两个复制叉。(b) 先进行强放射性脉冲标记、随后进行弱放射性脉冲标记后观察到的放射自显影图像的理想模式化图解，该模式图预测：如果成对条斑代表两个独立的单方向复制子，那么它们的复制叉移动的方向相同，后一种模式没有观察到。(c) 胚性蝾螈细胞复制 DNA 的放射自显影图像。成对条斑具有恒定的大小和形状，表明相应的复制子在同一时间开始复制。(*Sources*: (a) Huberman, J. A. and A. Tsai, Direction of DNA replication in mammalian cells. *Journal of Molecular Biology* 75 (1973) p. 8, by permission of Academic Press. (c) Callan, H. G., DNA replication in the chromosomes of eukaryotes. *Cold Spring Harbor Symposia on Quantitative Biology* 38 (1973) f. 4c, p. 195.)

示没有标记前体掺入的已完成复制的 DNA 区域。

图 20.11a 中成对条斑的形状是向外一端逐渐变细，十分类似于老式打蜡胡须。这表明在 DNA 复制过程中，先掺入了强放射性标记前体，然后再掺入了弱放射性标记前体，导致放射性强度从复制原点向两侧逐渐减弱。相反的实验过程即先进行“冷”标记，随后再进行“热”标记，则标记条斑的形状恰似倒转的打蜡胡须，向内逐渐变细。当然，这些成对条斑也可能是由空间距离相对较近的独立复制原点在进行复制时产生的，但是我们不可能指望这些复制原点总是以相反的方向进行复制。当然也有一些复制原点是以相同方向复制的，从而产生非对称性的放射自显影图像，如图 20.11b 中的假设。但事实上并没有观察到这种情况，所以，这些放射自显影实验的结果确证观察到的每对条斑确实只代表着一个复制原点，而不是相互靠近的两个复制原点。显然，果蝇的 DNA 是双向复制的。

这些实验结果都是以体外培养的成熟果蝇细胞为研究对象获得的，H. G. Callan 及同事用高放射性标记前体和两栖动物胚性细胞开展了类似的实验研究，如图 20.11c 所示，实验（以蝾螈胚性细胞为研究对象）结果是惊人的。与成熟昆虫细胞 DNA 复制的放射自显影图像相反，此实验获得的成对条斑的形状相同，长度大体一致，中间间距相等，表明这些复制原点是同时起始复制的，这一点是肯定的，因为掺入的标记前体出现在复制叉的相同位置处，

相对于各自复制原点的距离是相等的。此现象可能有助于解释蝾螈胚性细胞为何能够如此快速（仅需1h，而成熟细胞则需40h）地完成自身DNA的复制，所有复制原点能够同时而不是以错开的方式起始复制。

有关复制原点的讨论有助于我们阐释一个重要的术语——复制子（replicon）。一个DNA复制原点所控制的DNA序列称为一个复制子。因为只有一个复制原点，所以*E. coli*的整条染色体是一个复制子。显然，真核生物染色体包含有多个复制子，否则，完成整条染色体的复制需要花费太多的时间，即使以原核细胞每秒合成上千个核苷酸的DNA复制速度。事实上，真核生物DNA的复制速度要比原核生物DNA的复制速度慢得多。显而易见，真核生物染色体的复制需要多个复制子的参与。

单向复制

是否所有的遗传系统都采取双向复制的方式来完成DNA的复制？事实并非如此，如质粒colE1的复制。质粒是指是游离于细胞染色体之外能够独立复制的环状DNA分子（参见第4章），某些*E. coli*菌株含有colE1质粒，这也是该质粒名称中有“col”的原因。Lovett用电子显微镜对正在复制的colE1分子进行了观察，结果发现只有一个运动着的复制叉。

当然，如果实验者仅是简单地观察复制着的环形DNA分子，将只能观察到具有两个复制叉的复制泡，但无法确知复制是否发生在两个复制叉上。实验者需要DNA分子上拥有一些标记，这些标记能够在DNA复制期间的不同时间段内确定复制叉的位置。相对于标记，一方面，如果两个复制叉都发生了移动，说明两个复制叉都是活动着的，DNA进行双向复制；另一方面，如果只有一个复制叉移动，那么DNA进行单方向复制。

Lovett及同事选择的标记是colE1 DNA分子上的限制性内切核酸酶*Eco*RI的酶切位点。在第4章已经阐明*Eco*RI识别的序列为5′-GAATTC-3′，colE1 DNA分子上只有一个*Eco*RI识别位点，所以*Eco*RI能够将环形colE1 DNA切割形成线性分子，其末端产生于相同的位点即*Eco*RI酶切位点。图20.12显示了实验的结果，很明显，相对于线性DNA分子的末端，只有一个复制叉在移动，说明colE1 DNA是单向复制的。

小结 大多数真核生物与原核生物的DNA是进行双向复制的，colE1是DNA单向复制的一个事例。

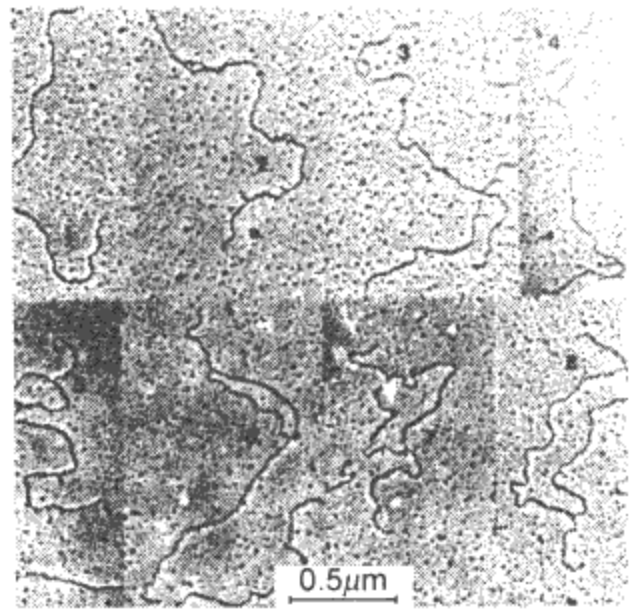


图20.12 colE1 DNA的单向复制。分子1还未开始复制，因此看不到复制泡的存在。分子2～8显示复制泡在逐渐增大，最值得注意的是：在复制泡的上方有一段没有复制的DNA，在各个图示中始终保持不变。相反，复制泡下方的一段DNA则随着复制泡增大而逐渐变短，很显然，只是下面的复制叉正在复制，而另一个复制叉处于静止状态。（*Source*：Lovett，M. A.，L. Katz and D. R. Helsinki，Unidirectional replication of plasmid ColE1 DNA. *Nature* 251（27 Sept 1974）p. 338. Macmillan Magazines Ltd.）

滚环复制

某些环形DNA分子采取滚环复制（rolling circle replication）的机制而不是θ模式进行复制。例如，ϕX174等具有单链环形基因组DNA的*E. coli*噬菌体采用相对简单的滚环复制机制完成自身DNA的复制，其双链复制型（replicative form，RFI）DNA分子能够产生多拷贝的单链子代DNA分子，图20.13简略概述了滚环复制机制。根据复制过程中间步骤（图20.13中的步骤b和c）的特点而将该复制模式称为滚环复制，正在复制的DNA分子的双链部分进行逆时针方向滚动，由此牵引出单链子代DNA分子，恰似展开的卫生纸卷在地板上迅速滚动。复制过程的中间分子类似于倒写的希腊字母σ（西格玛），所以滚环复制机制又称为σ模式，以区别于θ模式。

滚环复制并非只产生单链DNA分子，有些噬菌体（如λ噬菌体）能够利用滚环机制复制双链DNA分子。在λ DNA复制的早期阶段，噬菌体按θ复制模式产生若干个环状DNA分子拷贝，这些环状DNA分子并不被包装到噬菌体颗粒中，而是作为滚环复制的模板，合成能够被包装的线性λ DNA

分子。图 20.14 描述了滚环复制的发生过程。其复制叉十分类似于*E. coli* DNA 复制时的复制叉，前导链（绕环的那条链）的合成是连续进行，而后随链的合成是不连续的。在 λ 噬菌体中，子代 DNA 分子在被包装前会达到几个基因组的长度，这种多级长度的 DNA 分子称为连环体（concatemer）。包装机制只能将一个基因组长度的线性 DNA 分子包装到噬菌体的头部，所以，连环体必须在每一个完整基因组侧翼的 *cos* 位点处被酶切开。

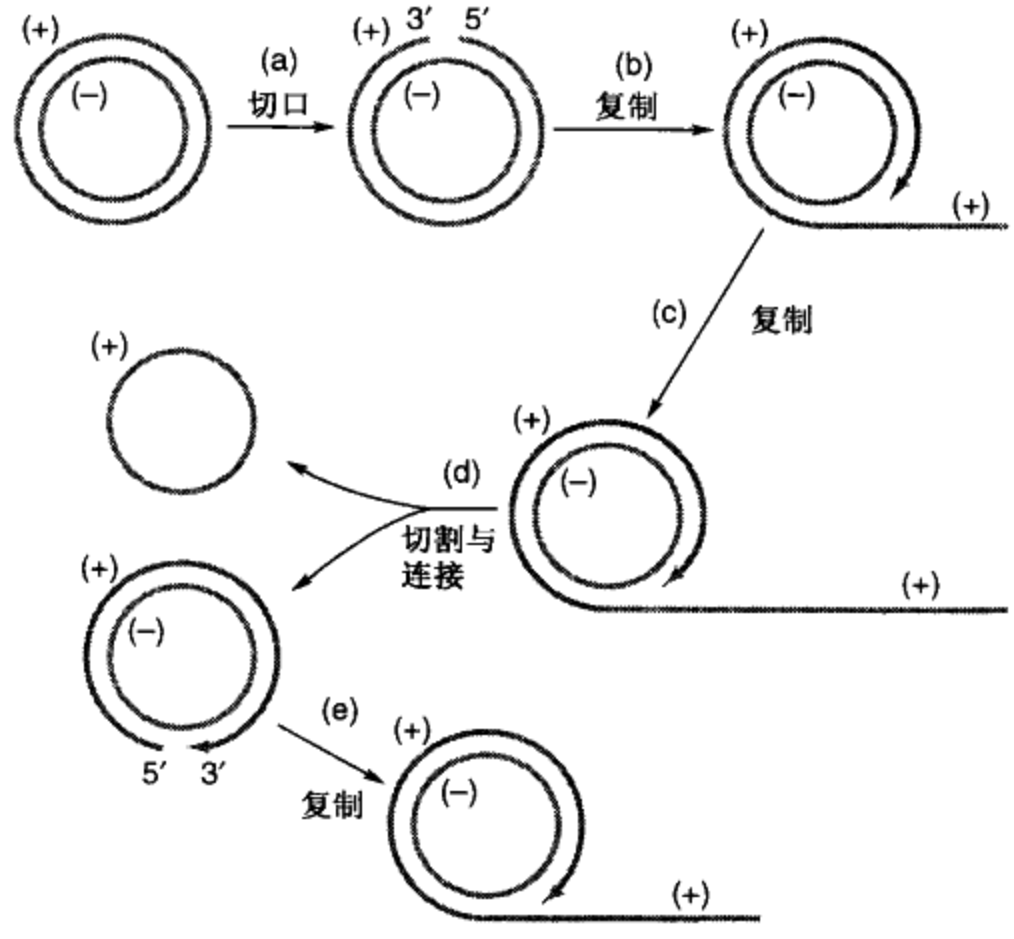


图 20.13　滚环复制产生单链环形子代 DNA 的示意图。（a）内切核酸酶在双链复制型 DNA 分子的正链上产生一个缺口；（b）缺口产生的游离 3′端成为正链延伸的引物，以负链为模板延伸正链的 3′端，从而将正链的 5′端置换出来。红色线条代表新合成的 DNA；（c）复制持续进行，使正链的长度接近于原长的 2 倍，DNA 环按逆时针方向滚动；（d）置换出的正链 DNA 达到单位长度时被内切核酸酶切离，然后环化；（e）继续以负链为模板复制合成新的正链。该过程重复发生，以此产生许多环形正链的拷贝。

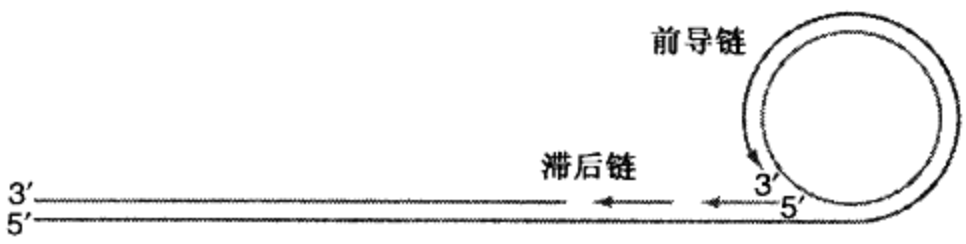


图 20.14　λ 噬菌体 DNA 的滚环复制模型。当环形 DNA 分子向右滚动时，前导链（红色）连续延伸，而后随链（蓝色）的延伸是不连续的。以开环的前导链为模板，利用 RNA 引物合成冈崎片段。在单个基因组长度的 DNA 被切下并包装到噬菌体头部之前，子代双链 DNA 延伸到多个基因组的长度（连环体）。

小结　环形 DNA 分子能够以滚环机制进行复制。双链 DNA 分子的一条被切割形成一个切口，以另一条完整的链为模板使切口的 3′端被延伸，从而将切口的 5′端置换出来。在噬菌体 ϕX174 的 DNA 复制过程中，当完成一轮复制后，就会释放出一个全长的单链环形 DNA 分子。在 λ 噬菌体中，置换链作为后随链进行不连续合成的模板。

20.2　DNA 复制的酶学

在*E. coli* DNA 复制过程中，至少需要 30 种不同多肽的协同运作。下面我们将分析其中某些蛋白质及其他生物体的同源蛋白的活性。

E. coli 中的 3 种 DNA 聚合酶

1958 年，Arthur Kornberg 发现了*E. coli* 的第一个 DNA 聚合酶。现在我们知道它只是 3 种 DNA 聚合酶中一种，称之为 DNA 聚合酶 I（DNA polymerase I，pol I）。在缺少能够证明细胞还存在其他 DNA 聚合酶证据的情况下，许多分子生物学家认为 pol I 负责细菌基因组的复制。正如我们将要了解到的那样，这个假设是不正确的，但是，我们还是从 pol I 开始对 DNA 聚合酶的学习，因为此酶相对简单且研究得较为透彻，同时该酶具有 DNA 合成酶的基本特征。

pol I　尽管 pol I 只是分子质量为 102kDa 的单链蛋白质，但却具有多种功能。pol I 能够催化 3 种

完全不同的反应，当然 pol I 具有 DNA 聚合酶活性，但还有 2 种不同的外切核酸酶活性：3′→5′外切核酸酶活性及 5′→3′外切核酸酶活性。为什么 DNA 聚合酶还需要拥有 2 种外切核酸酶活性？3′→5′ 外切核酸酶活性对于新合成 DNA 的校正是十分重要的（图 20.15）。如果 pol I 将错误的核苷酸添加到正在生长的 DNA 链上时，错误的核苷酸将不能与亲本链上的相应碱基进行正确的碱基配对，应该将这个错误核苷酸去除。于是，pol I 发生暂停，发挥其 3′→5′外切核酸酶活性将错配的碱基去除，从而使复制得以继续进行，并能够使 DNA 合成的忠实性或准确性大幅度提高。

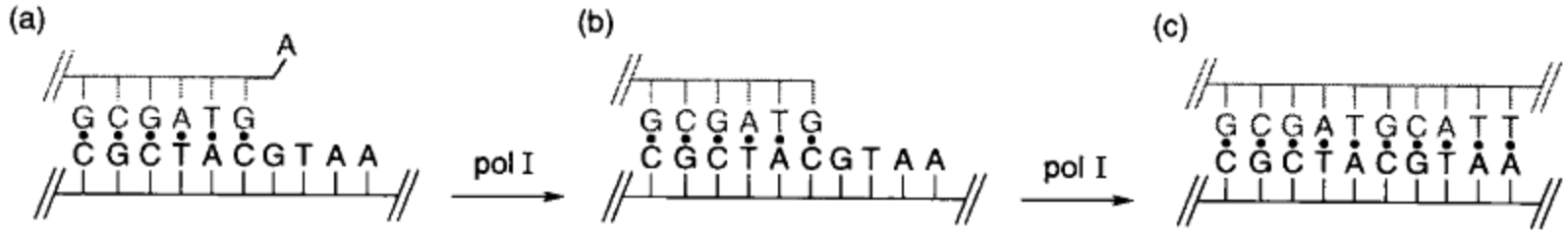


图 20.15　DNA 合成的校正。(a) 一个腺嘌呤核苷酸（红色）被错误地加入到胞嘧啶对面位置，破坏了引物 3′端所要求的碱基准确配对原则，造成复制的暂停；(b) 暂停后，pol I 发挥其 3′→5′ 外切核酸酶活性，将错配的核苷酸切除；(c) 当正确的碱基对得以修复后，pol I 继续合成 DNA。

5′→3′ 外切核酸酶活性使 pol I 能够降解聚合酶前方的 DNA 链，所以，当聚合酶经过时可以实现整条链的移除或置换，至少在体外是如此。由于 pol I聚合酶活性主要用于 DNA 的修复（包括 RNA 引物的去除和置换），因而其 DNA 降解功能显得十分重要，可以切除受损或错配的 DNA 序列（或 RNA 引物），以正确的 DNA 取而代之。图 20.16 描述了 RNA 引物去除与置换的过程。

pol I 还具有另一个重要的特征，在温和的蛋白酶水解作用下，该酶可被分割成 2 个肽段：较大片段称为 Klenow 片段（Klenow fragment），具有 DNA 聚合酶活性和 3′→5′外切核酸酶校正活性，较小片段具有 5′→3′外切核酸酶活性。在分子生物学研究中会经常用到 Klenow 片段，可用于 DNA 的合成、亲本链与引物的去除。例如，一方面，Klenow 片段经常用于 DNA 末端填充（参见第 5 章）及 DNA 测序；另一方面，在体外 pol I 全酶能够通过切口平移而标记探针，切口平移依赖于移动复制叉前方 DNA 的降解。

1987 年，Thomas Steitz 及同事测定了 Klenow 片段的晶体结构，首次观察到 DNA 合成系统的精细结构。该晶体结构的最显著特征是在 2 个 α 螺旋之间有一个裂隙，这可能是将要被复制的 DNA 结合位点。事实上，包括 T7 RNA 聚合酶在内的已知的所有聚合酶都具有相似的结构，并被形象地比作手形结构。在 Klenow 片段内，一个 α 螺旋是“手指”功能域的一部分，而另一个 α 螺旋则是“拇指”功能域的一部分，2 个 α 螺旋之间的 β 折叠是“手掌”功能域的一部分。“手掌”功能域包含有 3 个天冬氨酸残基，是酶发挥催化功能所必需的，与镁离子协同催化聚合反应。

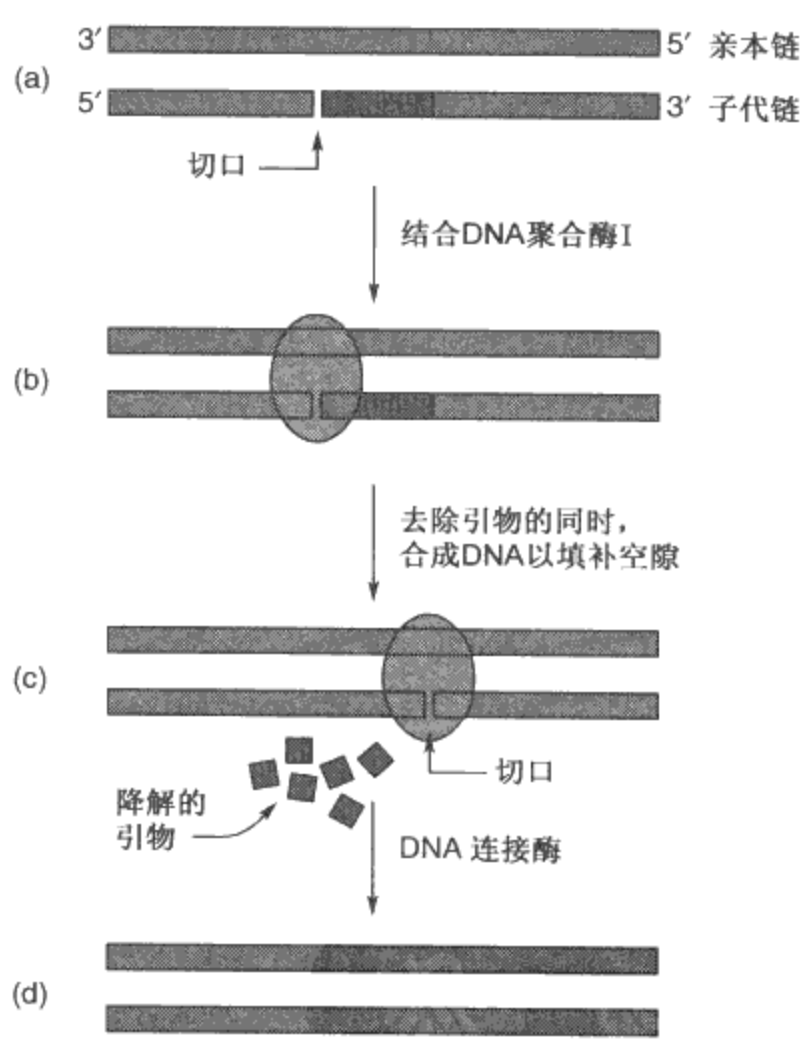


图 20.16　引物切除与新生 DNA 片段的连接。(a) 存在 2 个相邻的子代 DNA 片段，其中右侧片段的 5′端携带有 RNA 引物（红色），两片段被单链上的缺口分开；(b) DNA 聚合酶 I 在缺口处与双链 DNA 结合；(c) DNA 聚合酶 I 在去除引物的同时，通过向右延伸左侧片段而填补产生的空隙，引物也随之降解；(d) DNA 连接酶通过在左右子代 DNA 片段间形成磷酸二酯键，而将缺口缝合。

聚合酶结构中的裂隙是否是真正的 DNA 结合位点？为查明事实，Steitz 及同事获得了 *Taq* 聚合

酶的晶体以及长为 8bp 的双链 DNA 模板模型，其非模板（引物）链的 3′端为平末端。*Taq* 聚合酶是耐热细菌水生栖热菌（*Thermus aquaticus*）的 DNA 聚合酶，已被广泛应用于 PCR 反应（参见第 4 章）。它的聚合酶功能域因与 Klenow 片段的聚合功能域十分相似而被称为“KF”部分，即酶的 Klenow 片段部分。图 20.17 显示了 *Taq* 聚合酶-DNA 复合体的 X 射线晶体学研究结果，引物链（红色）的 3′端靠近手掌结构中 3 个重要的天冬氨酸残基，但并不十分靠拢，而是通过镁离子的桥接将天冬氨酸的羧基基团与引物链的 3′羟基基团联系起来，所以，该结构并不是一个精确的催化聚合反应的结构，部分原因可能是由于镁离子的缺失。

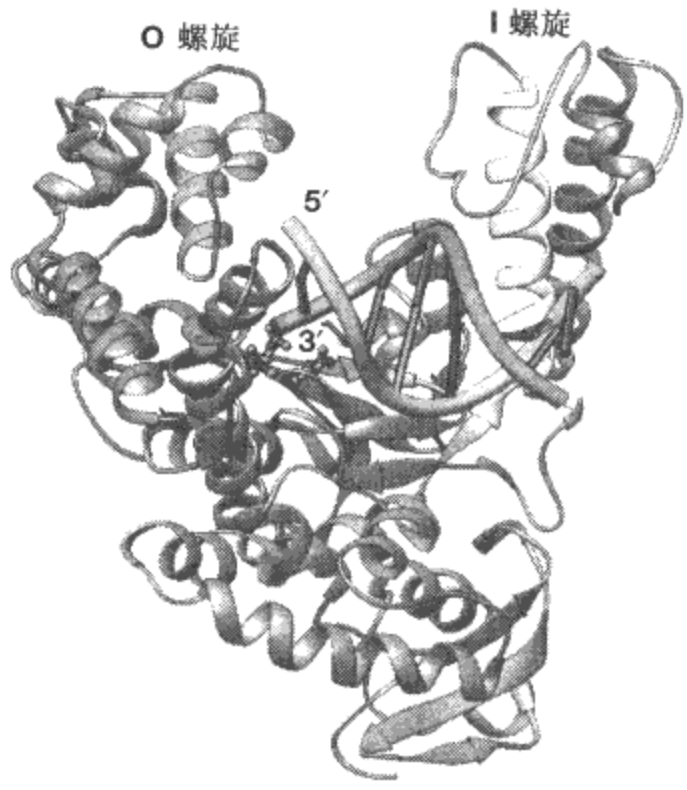


图 20.17　结合双链 DNA 模板的 *Taq* 聚合酶晶体结构。手形聚合酶“手指”和“拇指”的 O 螺旋和 I 螺旋分别用绿色和黄色标注，模式 DNA 的模板链和引物链分别用橙色和红色标注，手掌内 3 个必需天冬氨酸残基的侧链用红色小球体表示，靠近引物链的 3′端。（*Source*：Eom，S. H.，J. Wang and T. A. Steitz，Structure of *Taq* polymerase with DNA at the polymerase active site. *Nature* 382（18 July 1996）f. 2a，p. 280. Copyright © Macmillan Magazines，Ltd.）

1969 年，Paula DeLucia 和 John Cairns 分离获得了 *pol*A 基因缺陷突变体，该基因编码 pol I 蛋白。该突变体（*pol*A1）无 pol I 活性，但仍然能够存活，表明 pol I 并不是真正的 DNA 复制酶，其功能主要是进行 DNA 的损伤修复，当损伤的 DNA 被去除后，由 pol I 填补形成的空隙。pol I 不是主要 DNA 复制酶的发现驱使研究者重新寻找真正的 DNA 复制酶。1971 年，Thomas Kornberg 和 Malcolm Gefter 发现了聚合酶 II 和聚合酶 III（pol II 和 pol III），在后面的内容中，我们将会了解到 pol III 是真正的 DNA 复制酶。

小结　作为一种多功能酶，pol I 具有 3 种不同的功能：DNA 聚合酶活性、3′→5′外切核酸酶活性及 5′→3′外切核酸酶活性，酶的大功能域具有前两种活性，而 5′→3′外切核酸酶活性则存在于一独立的小功能域中。在温和蛋白酶水解作用下，大功能域（Klenow 片段）与小功能域分离，形成 2 个蛋白质片段，且仍然具有完整聚合酶所拥有的 3 种活性。Klenow 片段（包括所有已知的 DNA 聚合酶）的结构显示，该酶具有一个能够与 DNA 结合的裂隙，Klenow 片段的聚合酶活性位点与 3′→5′外切核酸酶活性位点相距较远。

pol II 和 pol III　通过磷酸纤维素层析法很容易将 pol II 与 pol I 区分开，但在野生型细胞内，因pol I 量多而不易检测到 pol III 的存在，如图 20.18 所示。Malcom Gefter、Thomas Kornberg 及同事用遗传学方法研究分析参与 DNA 复制的聚合酶，研究者在 15 个不同的温度敏感性*E. coli* 株系中测试了 pol II 和 pol III 的 DNA 复制活性，大部分株系为 *polA*1⁻ 突变体，因无 pol I 活性的竞争，经磷酸纤维素层析处理后容易测定 pol III 的活性。对于少数仍然具有 pol I 活性的菌株，Gefter 及同事用 *N*-乙烷基马来酰亚胺阻断 pol III 的活性，通过对比抑制剂存在与不存在两种情况下的活性差异，即可测定出 pol III 的活性。

表 20.1 显示了实验结果。最重要的发现是有 5 个株系的变异是 *dna*E 基因突变的结果，其中 4 个株系的 pol III 活性对温度十分敏感，而第 5 个株系的 pol III 活性对温度的敏感性不强。另外，所有突变株系中的 pol II 活性没有受到任何影响。根据这些实验结果可以推导出 3 个结论：第一，*dna*E 基因编码 pol III 蛋白；第二，*dna*E 基因不编码 pol II 蛋白，而且 pol II 和 pol III 具有独立的活性；第三，由于编码 pol III 蛋白的基因发生缺陷会妨碍 DNA 的复制，说明 pol III 是 DNA 复制所必需的。能够归结出 pol II 不是 DNA 复制所必需的结论，但根据本实验结果不可能获得这一结论，因为并没有测试编码 pol II 蛋白的基因发生缺陷的突变体。然而，在另一独立的实验中，研究者分离出了无 pol II 活性的突变体，这些突变体可以正常存活，表明 pol II 并不是 DNA 复制所必需的，所以，pol III 是 *E. coli* DNA 的复制酶。

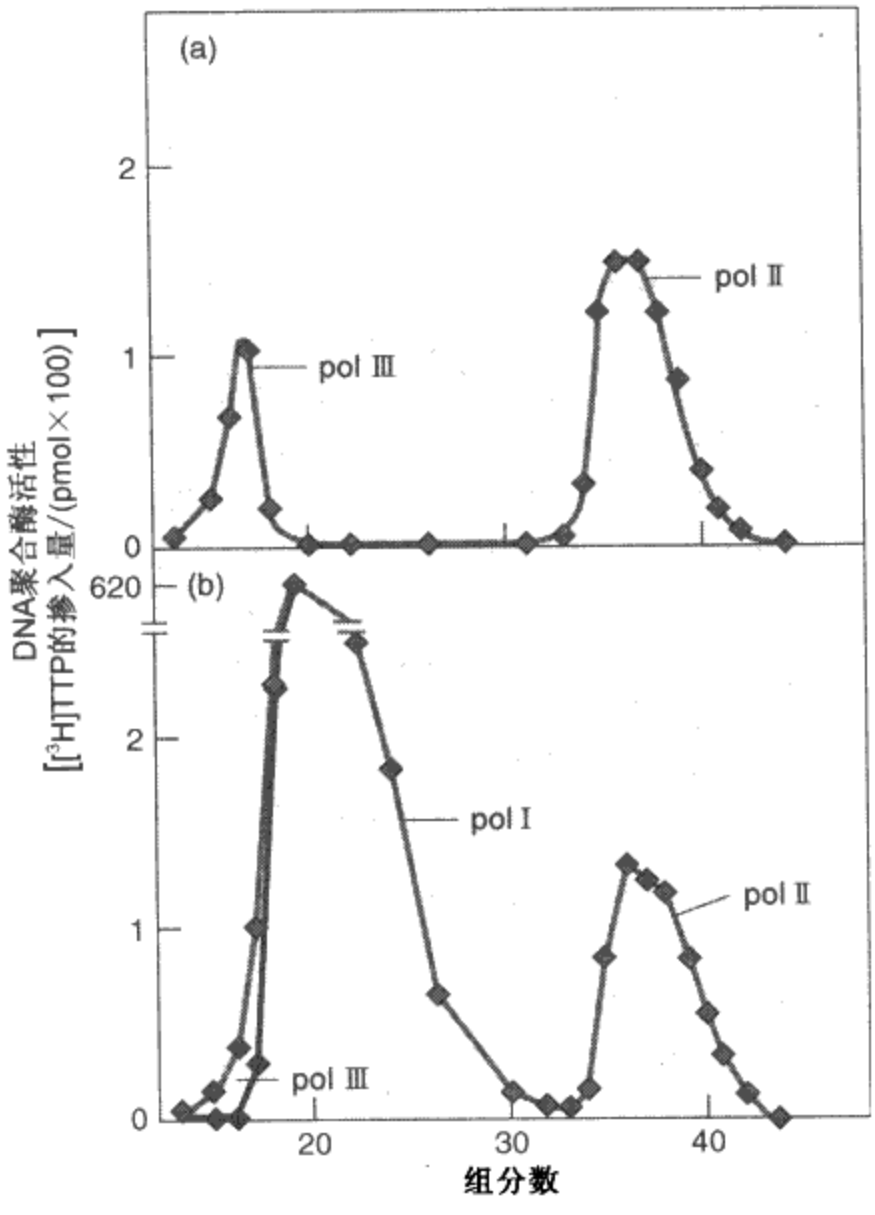


图 20.18　来自 *pol*A1 和野生型细胞中的 DNA 聚合酶的磷酸纤维素色谱分析。Gefter 及同事用磷酸纤维素膜色谱法分别从（a）*pol*A1 突变型（缺失 pol I）和（b）野生型*E. coli*细胞中进行 DNA 聚合酶的分离。pol II 易与 pol III（a）及 pol I（b）分开，但在野生型细胞中 pol III 被 pol I 掩盖（b）。在 *N*-乙烷基马来酰亚胺存在的情况下，通过对图（b）中组分的分析，即使在 pol I 存在的条件下也能检测到 pol III 的存在。*N*-乙烷基马来酰亚胺只抑制 pol III 而不抑制 pol I 的活性。因此，图（b）中红线与蓝线的差值即为 pol III 的活性。（*Source*：Adapted from Gefter, M. L., Y. Hirota, T. Kornberg, J. A. Wechster and C. Barnoux, Analysis of DNA polymerases II and III in mutants of *Escherichia coli* thermosensitive for DNA Synthesis, *Proceedings of the National Academy of Sciences USA* 68（12）: 3150-53, December 1971.）

表 20.1　在 45℃与 30℃两种温度条件下，DNA 聚合反应速度的比值

株系	DNA 聚合酶 II	DNA 聚合酶 III
1(*dnaA*)	1.85	1.45
2(*dnaB*)	1.73	1.74
3(*dnaB*)	1.95	1.55
4(*dnaC*)	1.86	1.20
5(*dnaD*)	1.99	1.47
6(*dnaE*)	1.87	1.0
7(*dnaE*)	1.70	—*
8(*dnaE*)	2.00	—*
9(*dnaE*)	1.95	<0.1
10(*dnaE*)	1.70	<0.1
11(*dnaF*)	1.55	1.33
12(*dnaF*)	1.77	1.43
13(*dnaG*)	1.75	1.57
14($dnaA^+$)	1.92	1.70
15($dnaA^+$)	1.90	1.55

* 破折号表示在 30℃温度条件下没有检测到酶的活性。

Source: From "Analysis of DNA Polymerases II and III in Mutants of *Escherichla coli*. Thermosensitive for DNA Synthesis," by M. L. Gefter, Y. Hirota, T. Kornberg, S. A. Wechster and C. Barnoux, *Proceedings of the National Academy of Sciences*, December 1971. Reprinted with permission of the author.

小结　在*E. coli*细胞的 pol I、pol II 和 pol III 这 3 种 DNA 聚合酶中，只有 pol III 是 DNA 复制所必需的，是细菌 DNA 的复制酶。

pol III 全酶　DNA 聚合酶 III 全酶（DNA polymerase III holoenzyme，pol III holoenzyme）通过延伸引物合成 DNA 的前导链和后随链，之所以称之为“全酶”是因为 DNA 聚合酶 III 由多亚基组成，如表 20.2 所示，DNA 聚合酶 III 全酶包含有 10 个不同的多肽，稀释后，全酶解离成几个不同的亚聚体（见表 20.2）。每一个 pol III 亚聚体均具有 DNA 聚合能力，但聚合速度很慢，而在体内 DNA 复制的速度很快，这表明亚聚体缺失了某些重要组分。*E. coli* 的 DNA 复制叉能够以 1000nt/s 的惊人速度移动（设想一下，该系统能够在 1s 时间内使 1000 个核苷酸与亲本链上的碱基进行准确配对，形成 1000 个磷酸二酯键）。在体外，DNA 聚合酶 III 全酶能够以 700nt/s 的速度合成 DNA，接近于体内复制的速度，表明 DNA 聚合酶 III 全酶就是体内的 DNA 复制系统。细胞内的另两类 DNA 聚合酶 pol I 和 pol II 一般不以全酶的形式存在，且 DNA 复制速度要远小于 pol III 全酶的复制速度。

表 20.2 *E. coli* DNA 聚合酶 III 全酶的亚基组成

亚基	分子质量/kDa	功能			
α	129.9	DNA 聚合酶	核心	pol III′	pol III 全酶
ε	27.5	3′→5′外切核酸酶			
θ	8.6	促进 ε 的外切酶活性			
τ	71.1	结合二聚体核心与 γ 复合体			
γ	47.5	结合 ATP	γ 复合体（DNA 依赖的 ATP 酶）	pol III*	
δ	38.7	结合 β			
δ′	36.9	结合 γ 和 δ			
χ	16.6	结合 SSB			
Ψ	15.2	结合 χ 和 γ			
β	40.6	滑动钳			

Source: Reprinted from Herendee, D. R. and T. T. Kelly, DNA Polymerase III: Running rings around the fork *Cell* 84: 6, 1996. Copyright © 1996, with permission from Elsevier.

Charles McHenry 和 Weldon Crow 纯化获得了近乎同质的 pol III，其中 pol III 核心由 3 条多肽组成，即 α、ε 和 θ 亚基，分子质量分别为 130kDa、27.5kDa 和 10kDa。在纯化过程中其余的亚基已发生了解离，但这 3 个亚基却紧密地结合在一起。本节将详尽分析 pol III 核心的功能特征，将在第 21 章中对 pol III 全酶的其他亚基进行探讨，这些亚基在 DNA 合成的起始与延伸过程中发挥着重要功能。

pol III 核心酶中的 α 亚基具有 DNA 聚合酶活性，因不易与其他核心亚基分离，所以很难测定 α 亚基的 DNA 聚合酶活性。当 Hisaji Maki 和 Arthur Kornberg 克隆了 α 亚基的基因并使之超表达后，研究者最终提纯了具有 DNA 聚合酶活性的 α 亚基，通过基因的超表达而使 α 亚基的数量远远多于其他 2 个亚基的数量。在测试纯化 α 亚基的 DNA 聚合酶活性时，研究者发现 α 亚基所具有的活性相当于等量核心酶所具有的活性，因此可以断定是 α 亚基赋予了核心酶的 DNA 聚合酶活性。

pol III 核心酶具有 3′→5′外切核酸酶活性，能够将切除掺入的错配碱基，从而使聚合酶在 DNA 合成过程中具有校正功能，类似于 pol I Klenow 片段的 3′→5′ 外切核酸酶活性。Scheuermann 和 Echols 运用超表达策略论证了核心亚基 ε 具有 3′→5′ 外切核酸酶活性。研究者过量表达 ε 亚基（*dna*Q 基因的编码产物）后，并通过不同步骤将其纯化，最后经 DEAE-Sephacel 离子交换层析而获得了较为纯净的 ε 亚基（泳道 D）。随后，Richard Scheuermann 和 Harrison Echols 测试了纯化的 ε 亚基以及 pol III 核心的外切酶活性。图 20.19 表明 pol III 核心与 ε 亚基均具有外切核酸酶活性，对错配的 DNA 底物具有专一性，但对准确配对的 DNA 则无作用，这正是所预期的校正活性。这种活性也能够解释 *dna*Q 基因突变会引起 DNA 的大量变异（超过野生型细胞的 10^3～10^5 倍），若没有经过充分的校正，许多错配的碱基就不能被切除，从而导致突变。将 *dna*Q 突变体称为增变突变体（mutator mutant），因增变表型的存在而将该基因称为 *mut*D 基因。

对核心酶中的 θ 亚基的研究工作相对较少，除了能够促进 ε 亚基的外切核酸酶活性外，其他的功能还是未知的。但有一点是清楚：在核心酶中 α 亚基和 ε 亚基能够彼此增强活性。与游离状态相比，α 亚基的 DNA 聚合酶活性升高了 2 倍，ε 亚基的外切核酸酶活性升高了 10～80 倍。

小结 pol III 核心酶由 α、ε 和 θ 3 个亚基组成。α 亚基具有 DNA 聚合酶活性，ε 亚基具有 3′→5′ 外切核酸酶活性，能够执行校正功能，θ 亚基的作用目前还不清楚。

复制保真性

pol III（和 pol I）的校正机制极大地提高了 DNA 复制保真性。在体外，pol III 核心酶大约每合成 10 万个碱基对会出现一个错误，复制的准确率并不理想，因为*E. coli* 的基因组包含 400 万个碱基对，按这样的错误概率计算，每繁殖一代，通过 DNA 复制就会导致基因产生较大比例的错误。幸运的是，聚合酶的校正机制能够确保碱基准确配对，校正机制的出错率可能与复制的错误率相同，大约均为 10^{-5}，由此可以推测，经过校正后，DNA 复制的出错率仅为 $10^{-5}\times10^{-5}=10^{-10}$，接近于 pol III 全酶在体内进行 DNA 复制时 10^{-10}～10^{-11} 的出错率（复制保真性至少部分得益于错配修复，相关内容将在本

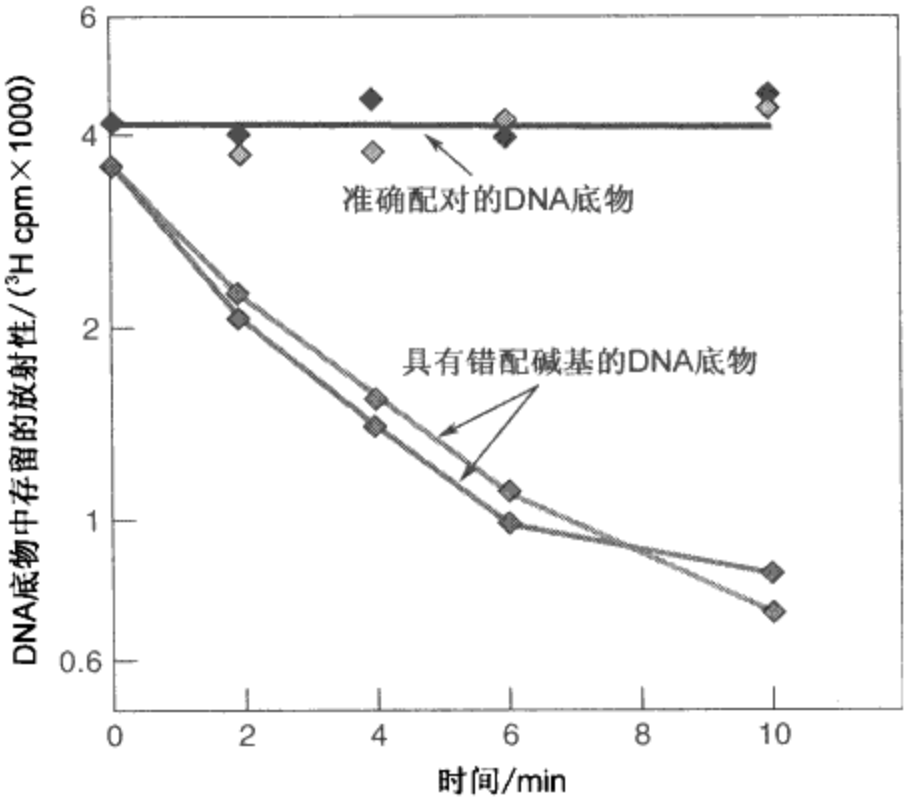


图 20.19　ε 亚基和 pol III 核心酶对准确配对或错配 DNA 底物的外切核酸酶活性。Scheuermann 和 Echols 将纯化的 ε 亚基与^{3}H 标记的人工合成 DNA 产物共同温育。在逐渐延长温育时间的情况下，测定温育后 DNA 底物中留存的放射性。蓝色与绿色为 pol III 核心酶作用的结果，橙黄色与红色为 ε 亚基作用的结果。(*Source*: Adapted from Scheuermann, R. H. and H. Echols, A separate editing exonuclease for DNA replication: The ε subunit of *Escherichia coli* DNA polymerase III holoenzyme. *Proceedings of the National Academy of Sciences USA* 81: 7747-51, December 1984.)

章后面讨论)。事实上，生物体可以忍受这种程度的保真性，这比复制的绝对保真要好，通过复制产生适度变异，其中某些变异可以使生物体更好地适应变化了的环境。

校正机制的功能是从 DNA 子链的 3′端将错配的核苷酸切除（图 20.15）。如果没有已配对的核苷酸提供 3′端，使 DNA 聚合酶遵循碱基互补配对原则将游离的核苷酸添加到 3′端，那么 DNA 聚合酶就无法起始新 DNA 链的合成，除非在引物存在的情况下，DNA 聚合酶才能起始新 DNA 链的合成。这就解释 DNA 复制需要引物的原因，但为什么一定是 RNA 引物？其可能的原因是：合成引物时，因未经过校正而具有较高的出错率，RNA 引物易被识别而切除，并通过冈崎片段的延伸而被 DNA 所取代，当然，相对而言，后一个过程一般是无差错的，因为该过程由具有校正功能的 pol I 催化完成。

小结　DNA 复制的保真性对生命而言是至关重要的，为确保 DNA 复制的保真性，*E. coli* 的 DNA 复制机构具有一个内在的需要引发的校正系统。只有配对的核苷酸才能作为 pol III 全酶起始 DNA 复制的引物。因此，如果偶尔掺入了错误的核苷酸，复制将会出现暂停，直至 pol III 的 3′→5′ 外切核酸酶活性将其切除。而 RNA 引物可成为其降解的标签。

真核生物的多种 DNA 聚合酶

对参与真核生物 DNA 复制的蛋白质我们还知之甚少，但已经知道有多种 DNA 聚合酶参与了 DNA 复制过程，并对这些酶的作用有了很好的了解。表 20.3 列举了哺乳动物主要的 DNA 聚合酶及其可能的作用。

表 20.3　某些真核 DNA 聚合酶可能的功能

酶	可能具有的功能
DNA 聚合酶 α	引发两条链的复制
DNA 聚合酶 δ	延伸两条 DNA 链
DNA 聚合酶 β	DNA 修复
DNA 聚合酶 ε	DNA 修复
DNA 聚合酶 γ	线粒体 DNA 的复制

由于 DNA 聚合酶的进行性较低，过去曾认为 DNA 聚合酶 α 负责后随链的合成。进行性（processivity）是指聚合酶一旦起始 DNA 复制后所能持续进行的时间或聚合 DNA 链的长度，*E. coli* pol III 全酶具有很高的进行性，一旦起始 DNA 链的合成，pol III 全酶就能够长久地与 DNA 模板保持结合状态并持续进行 DNA 的合成。*E. coli* DNA 复制的整体速度非常快，因为聚合酶不会频繁地与模板解离，如果发生解离，复制就会出现暂停，需要新的聚合酶结合上去并重新开始复制。与聚合酶 α 相比，聚合酶 δ 拥有更高的进行性，于是有研究者认为进行性较低的聚合酶 α 负责由短片段构成的后随链的合成，然而，作为真核生物唯一具有引发酶活性的 DNA 聚合酶 α，聚合酶 α 能够为两条链合成引物，然后由高进行性的 DNA 聚合酶 δ 完成两条链的延伸。

事实上，聚合酶 δ 的高进行性不是源于酶自身，而是由于**辅助蛋白增殖细胞核抗原**（proliferating cell nuclear antigen，**PCNA**）的作用。该蛋白质在 DNA 复制活跃的增殖细胞中富集，能够使聚合酶 δ 的进行性提高 40 倍，即与单独延伸 DNA 的长度相比，PCNA 能够使聚合酶 δ 在脱离模板前将 DNA 链

延长了 40 倍。PCNA 的作用机制是将聚合酶 δ 钳在 DNA 模板上，在第 21 章学习*E. coli* DNA 复制的详细机制时，我们再更加全面地研究这种钳夹现象。

与高进行性的聚合酶 δ 形成鲜明对比的是聚合酶 β 根本就不具有进行性，通常在生长的 DNA 链上添加一个核苷酸后就与模板解离，因此需要一个新的聚合酶 β 结合上来并完成下一个核苷酸的添加。这与此酶可能具有的修复功能相吻合，当引物或错配的碱基被切除后，需要聚合酶 β 合成短的 DNA 片段以填补形成的空隙。此外，聚合酶 β 在细胞内的水平并不受细胞分裂速率的影响，表明它不参与 DNA 的复制。如果聚合酶 β 参与 DNA 复制，则会像聚合酶 δ 和 α 一样，在快速分裂的细胞内其含量应该会显著增多。

聚合酶 γ 存在于线粒体而不是细胞核内，负责线粒体 DNA 的复制。

小结 哺乳动物细胞内存在 5 种不同的 DNA 聚合酶。聚合酶 δ 和 α 参与 DNA 两条链的复制，聚合酶 α 引发 DNA 的复制，而聚合酶 δ 负责两条链的延伸。聚合酶 β 和 ε 似乎是在 DNA 修复中起作用，聚合酶 γ 可能负责线粒体 DNA 的复制。

链的分离

在讨论 DNA 复制的一般特征时，我们假定 DNA 的两条链在复制叉处以某种方式发生了分离。当 DNA 聚合酶在进行 DNA 复制时，两条彼此紧密结合的亲本链不会自动发生解离，需要能量及酶的作用才能使两条亲本链发生解离。

解旋酶 利用 ATP 化学能在复制叉处将两条 DNA 亲本链发生解离的酶称为解旋酶（helicase）。在第 17 章我们讨论真核生物翻译起始过程时，已经学习了有关解旋酶的一个事例，但那是 RNA 解旋酶，能够使 mRNA 前导区的发夹结构发生解旋，而 DNA 解旋酶则是将复制叉前方的 DNA 链解离。

已从 *E. coli* 细胞中鉴定出了多种 DNA 解旋酶，但关键的问题是要确定哪一个解旋酶参与了 DNA 的复制。在所研究的解旋酶中，前 3 个酶，即 *rep* 解旋酶、DNA 解旋酶 II 和 DNA 解旋酶 III 发生突变后均不会影响细胞的增殖，说明这 3 种酶不可能参与像 DNA 复制这样涉及细胞生死存亡的重要生命过程。可以预料，如果参与 DNA 复制过程的解旋酶出现缺陷，则会产生致死突变效应。

获得关键基因缺陷突变的途径之一是进行条件突变，通常是温度敏感型突变。温度敏感型突变体的特点是：在低温条件下培养的突变细胞不会表现出突变表型，若升高温度，突变细胞即可表现出突变表型。早在 1968 年，François Jacob 及同事就发现了两类与*E. coli* DNA 复制有关的温度敏感型突变体（图 20.20），当温度从 30℃升高到 40℃时，1 型温度敏感型突变体会立即关闭 DNA 的合成，而 2 型温度敏感型突变体在提升温度的条件下，会逐渐降低 DNA 合成的速率。

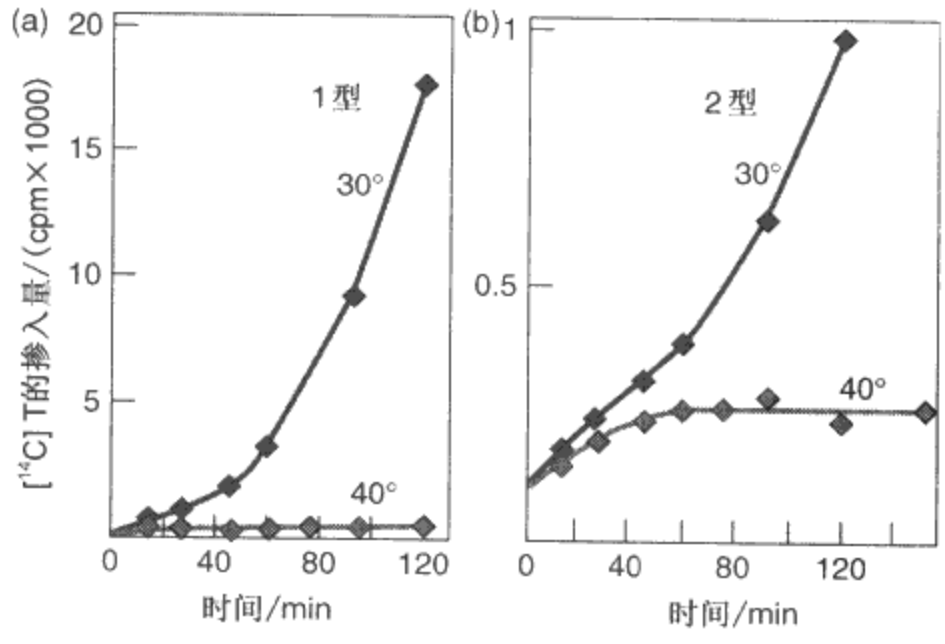


图 20.20 两类*E. coli*DNA 复制突变体。Jacob 及同事对 DNA 复制的温度敏感型突变体进行测试。他们将一半培养物转移到非允许温度（40℃）条件下培养，另一半在允许温度（30℃）条件下培养，然后加入[^{14}C]胸苷标记新合成的 DNA。鉴别出两类突变体：(a) 1 型突变体，当温度从 30℃升高到 40℃时，1 型温度敏感型突变体会立即关闭 DNA 的合成；(b) 2 型突变体，在非允许温度条件下，DNA 合成速率逐步下降。（*Source*: Adapted from Hirota, Y., A. Ryter and F. Jacob, Thermosensitive mutants of *E. coli* affected in the processes of DNA synthesis and cellular division, *Cold Spring Harbor Symposia on Quantitative Biology* 33: 678, 1968.）

其中有一种 1 型突变体为 *dna*B 突变体，*E. coli* DNA 合成温度敏感型 *dna*B 突变体细胞在温度升高到非允许水平时，DNA 的合成即可发生停止，与预期结果相一致，即如果 *dna*B 基因编码的 DNA 解旋酶确实是 DNA 复制过程所必需的，那么当该功能性解旋酶缺失时，复制叉将停止移动，DNA 的合成也会立即停止。此外，*dna*B 的编码产物（DnaB）是一种 ATP 酶，这也是预期的 DNA 解旋酶应具有的活性，dnaB 蛋白与合成 DNA 复制引物的引发酶相关。

所有的这些研究结果表明，DnaB 作为解旋酶，能够在*E. coli* DNA 复制过程中使 DNA 双螺旋解链。为证实 DnaB 确实具有 DNA 解旋酶活性，Jonathan LeBowitz 和 Roger McMacken 于 1986 年开展了这项研究工作，实验所用的解旋酶底物如图 20.21a 所示，该底物为环形 M13 噬菌体 DNA，退火结合有一短的线性 DNA 片段，该片段的 5′端被标记。图 20.21a 同时也显示了解旋酶分析实验的操作过程。LeBowitz 和 McMacken 将标记底物与 DnaB 或其他蛋白质共同温育，然后对反应产物进行电泳分析，如果反应体系中的蛋白质具有解旋酶活性，那么该蛋白质就能使双螺旋 DNA 解旋而形成两条单链，进行电泳时，短的标记线性 DNA 具有较高的电泳迁移率，在凝胶中将与未标记的大分子 DNA 进行独立迁移。

图 20.21b 显示分析实验的结果。DnaB 自身具有解旋酶活性，并受 DnaG（我们将在第 21 章介绍的引物酶）和单链 DNA 结合蛋白 SSB 的促进，有关 SSB 的内容将在下节讲述。单独的或结合在一起的 DnaG 和 SSB 不具有 DNA 解旋酶活性，所以 DnaB 是在复制叉处实现 DNA 解链的解旋酶。

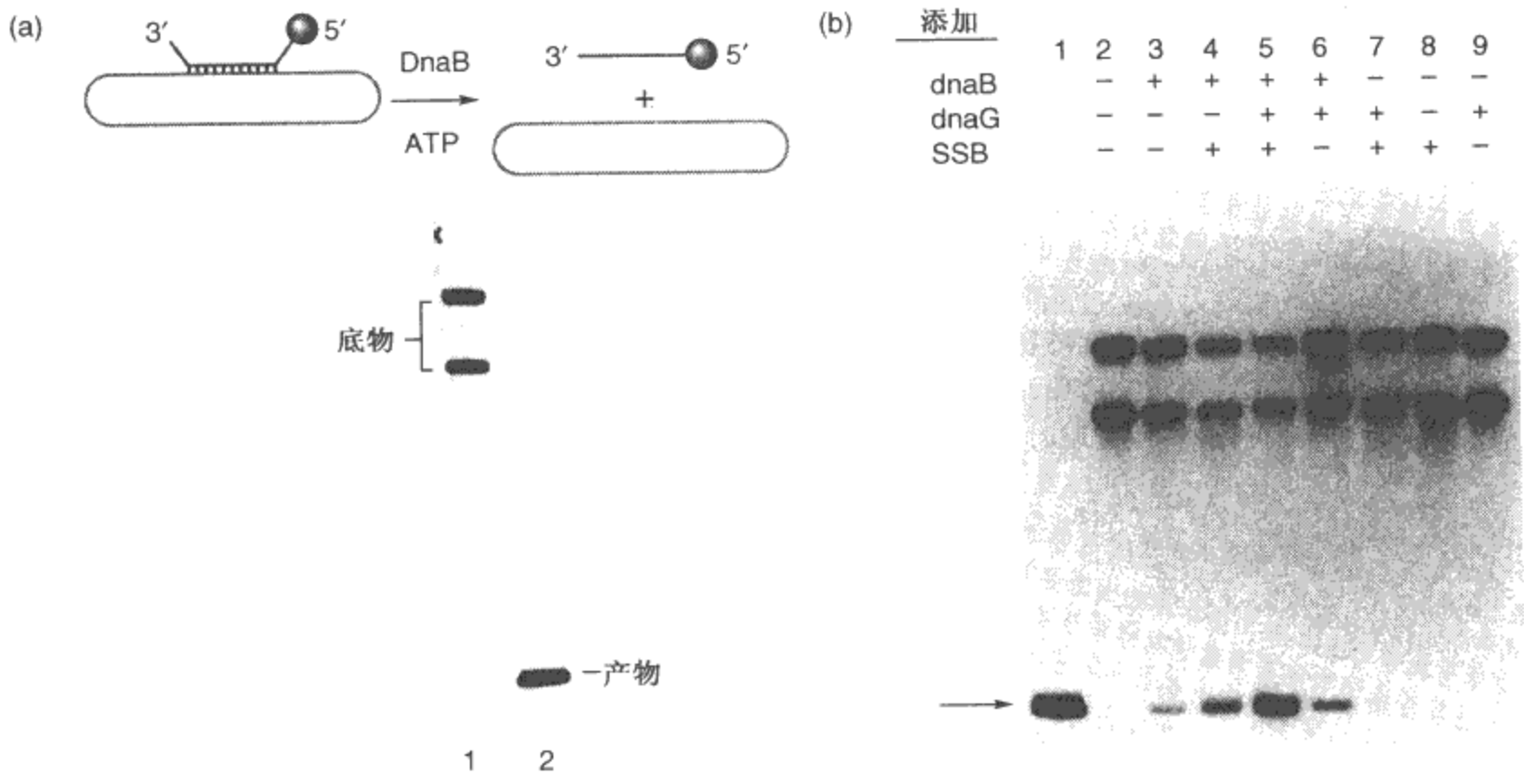


图 20.21　DNA 解旋酶分析。（a）分析的原理。LeBowitz 和 Roger McMacken 合成了解旋酶的底物（图上方），用^{32}P 标记长度为 1.06kb 的单链线性 DNA（红色）的 5′端，使之与含有 1.06kb 线性 DNA 互补区域、未被标记的重组单链 M13 噬菌体 DNA 退火，DnaB 蛋白或任何其他的 DNA 解旋酶能够将底物的双链区解旋，从而使短的线性标记 DNA 片段（红色）释放出来。图板底部：底物凝胶电泳后产生 2 条带（泳道 1），可能分别对应于与短的线性标记 DNA 退火的线性及环形较长的 DNA 链。将短的 DNA 片段单独电泳（泳道 2）显示，与底物相比较，具有较高的迁移率（标有“产物”的条带）。（b）解旋酶分析结果。LeBowitz 和 Roger McMacken 在反应体系中加入 DnaB、DnaG 和 SSB（图上方标注）后，按图（a）实验方案进行实验分析，图的底部为凝胶电泳结果。泳道 1 为未经退火的短标记 DNA 片段的电泳结果（箭头），泳道 3 的结果表明 DnaB 自身具有解旋酶活性，泳道 4 和 5 表明其他蛋白质能够促进 DnaB 的解旋酶活性。另一方面，泳道 7～9 说明没有 DnaB 时，DnaG 和 SSB 这两种蛋白质均没有 DNA 解旋酶活性。（*Source*：LeBowitz, J. H. and R. McMacken, The *Escherichia coli* dnaB replication protein is a DNA helicase. *Journal of Biological Chemistry* 261 (5 April 1986) figs. 2, 3, pp. 4740-41. American Society for Biochemistry and Molecular Biology.）

小结 在复制叉处将双链 DNA 解链的解旋酶是由*E. coli*的*dnaB* 基因所编码。

单链 DNA 结合蛋白

正如已了解的那样，在复制过程中，单链 DNA 结合蛋白（single strand DNA-binding proteins，SSB）也参与了 DNA 双链的解离。这类蛋白质并不像解旋酶那样能够催化 DNA 链的解离，但是一旦有单链 DNA 形成，SSB 便会选择性的立即与单链 DNA 结合，覆盖在单链 DNA 上的 SSB 能够阻止单链 DNA 重新缔合成双螺旋，单链 DNA 的形成主要是正常 DNA 呼吸作用（DNA 分子在局部发生链的瞬间分离，特别在富含 A-T 碱基对的区域）或解旋酶作用的结果，SSB 的结合有助于维持单链状态。

对原核生物 SSB 的研究最为深入。*E. coli* 的 SSB 由 *ssb* 基因所编码；T4 噬菌体的单链 DNA 结合蛋白是 gp32，意为“基因产物 32”（gene product 32，即 T4 噬菌体基因 32 编码的产物）；M13 噬菌体的单链 DNA 结合蛋白是 gp5（M13 噬菌体基因 5 的编码产物）。这些蛋白质都有协同效应：一个 SSB 的结合会促进下一个 SSB 与单链 DNA 的结合。例如，第一个 gp32 蛋白分子与单链 DNA 的结合，则会使下一个 gp32 蛋白分子与单链 DNA 的亲和力提升上 1000 倍，所以，一旦第一个 gp32 蛋白分子完成与单链 DNA 的结合，那么，第二个 gp32 蛋白分子与单链 DNA 的结合就会变得很容易，第三更是如此，依此类推。从而单链 DNA 区域被成串的 gp32 分子所覆盖。gp32 分子链甚至可以伸展到双链的发夹结构内，只要 gp32 协同结合所释放的自由能多于发夹结构形成时所释放的自由能，gp32 分子链就可使发夹结构发生解链。实际上，相对较小、碱基配对不理想的发夹结构易被解链，而相对较长、碱基配对完全匹配的发夹结构会仍然保持完整状态。gp32 蛋白以单体形式与单链 DNA 结合，而 gp5 是以二聚体形式，*E. coli* 的 SSB 以三聚体形式与单链 DNA 结合。

目前，我们已认为“单链 DNA 结合蛋白”的称谓会引起一定程度的误解，这些蛋白质确实能够与单链 DNA 结合，但在前面的章节中了解到有许多蛋白质都可以与单链 DNA 结合，包括 RNA 聚合酶，只不过是 SSB 与单链 DNA 结合的较为频繁而已。SSB 不仅能够使 DNA 保持单链状态，同时也能特异性地促进其同源蛋白 DNA 聚合酶的作用。例如，gp32 蛋白能够促进 T4 DNA 聚合酶的作用，但对 T7 聚合酶或*E. coli* DNA 聚合酶 I 则无促进作用。

SSB 的活性是否具有重要性？事实上，SSB 为必需蛋白质，在非许可温度条件下，*E. coli ssb* 基因温度敏感性突变体细胞是致死的，如果该突变体细胞被 T4 噬菌体的 gp32 温度敏感型突变体 tsP7 侵染后，在转移到非许可温度条件下，2min 内噬菌体 DNA 的复制就会停止（图 20.22）。而且，噬菌体 DNA 开始被降解。该特征表明 gp32 蛋白的功能之一是防止在噬菌体复制过程中形成的单链 DNA 被降解。

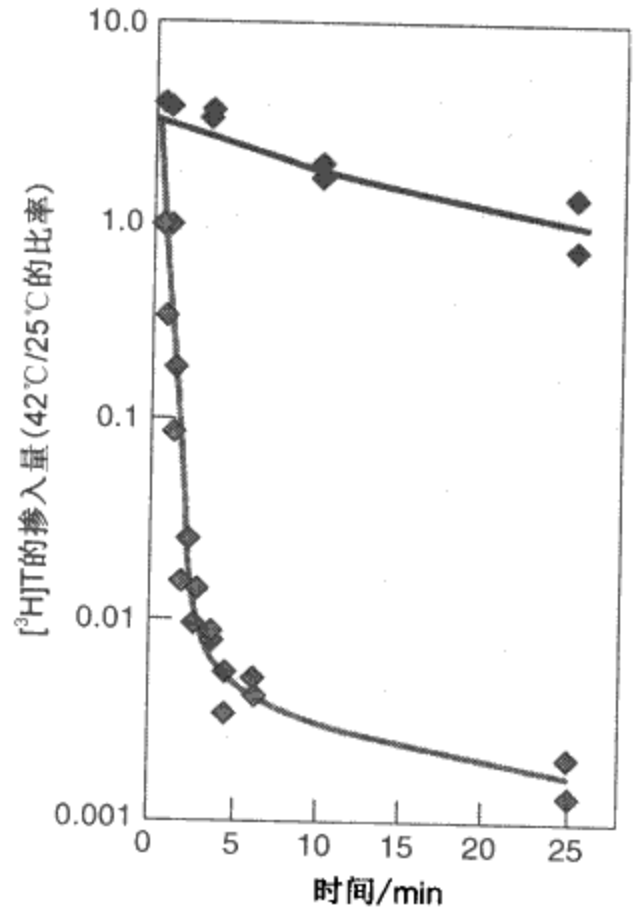


图 20.22 SSB（gp32）基因温度敏感型 T4 突变体侵染后，细胞 DNA 合成的温度敏感性。在被 T4 噬菌体突变体侵染后，于 42℃和 25℃条件下脉冲标记 1min 后，Curtis 和 Alberts 测定了［^{3}H］胸苷的相对掺入量。T4 噬菌体突变体分别为：基因 23（蓝色）；基因 32 及基因 23（红色）；基因 32 及基因 49（黄色）。基因 23 和 49 的琥珀型突变并不影响 DNA 的合成，因此，DNA 合成的下降是基因 32 发生了温度敏感型突变所引起的。（*Source*：Adapted from Curtis，M. J. and B. Alberts，Studies on the structure of intracellular bacteriophage T4 DNA，*Journal of Molecular Biology*，102：793-816，1976.）

SSB 在原核生物中具有如此重要的作用，但令人奇怪的是，在真核生物中并没有发现具有相似重要性的 SSB。然而在人类细胞中发现，宿主的 SSB 或 RF-A 蛋白是 SV40 DNA 复制所必需的，该蛋白质能选择性地与单链 DNA 结合，并促进病毒大 T 抗原的 DNA 解旋酶活性。由于 RF-A 为宿主蛋白，

我们推测该蛋白质在未被病毒侵染的人类细胞内发挥一定的作用，只是我们还不知道是什么作用。病毒基因编码的 SSB 在某些真核生物病毒 DNA 复制过程中发挥主要作用，包括腺病毒和疱疹病毒的 DNA 复制。

小结 原核生物单链 DNA 结合蛋白与单链 DNA 的结合力远强于与双链 DNA 的结合力。单链 DNA 结合蛋白以协同方式能够与新形成的单链 DNA 进行紧密结合，从而协助解旋酶发挥作用。覆盖在单链 DNA 上的 SSB 能够阻止单链间的退火，而且能够起保护作用，防止单链 DNA 被降解，此外，SSB 对同源的 DNA 聚合酶还有促进作用。SSB 的这些活性对于原核生物 DNA 复制是非常重要的。

拓扑异构酶

有时我们将 DNA 链的分离称之为“解拉链”，在使用这一术语时不应忘记，DNA 分子并不像拉链那样两侧平行且平直，而是呈螺旋结构。因此，DNA 的两条链必须通过彼此间相互旋转才能实现分离，如果 DNA 分子是线性结构且很短，那么解旋酶自身即可实现 DNA 的解链，但在对闭合环形 DNA 分子如 *E. coil* 染色体解旋时则会出现一个特殊的问题，当 DNA 在复制叉处解旋时，在环形 DNA 分子的其他区域就会产生补偿性缠绕，这种因螺旋紧密缠绕而产生的张力必须获得释放，否则 DNA 双螺旋分子无法承受这种张力。1963 年，Cairns 在首次观察 *E. coli* 环形 DNA 分子时认识到该问题的存在。他提出，在 DNA 双螺旋分子上存在一个旋转体，能够使 DNA 分子向相反的方向旋转，从而使因解旋产生的张力得以释放（图 20.23）。现在已经知道是 DNA 旋转酶（DNA gyrase）承担旋转作用，DNA 旋转酶属于拓扑异构酶（topoisomerase），能够使 DNA 单链或双链发生瞬间断裂而改变其拓扑结构。

为了解拓扑异构酶的作用机制，我们需要仔细分析在第 2 章和第 6 章学习到的 DNA 超螺旋现象。到目前为止，已被研究的所有自然形成的闭合环形双链 DNA 分子都是以超螺旋形式存在的，闭合环形 DNA 是指无单链断裂或具缺口的 DNA 分子，当细胞产生闭合环形超螺旋 DNA 分子时，必然会引起双螺旋发生某种程度的解旋，这时 DNA 分子处于解旋状态。只要两条链保持完好，就不会发生以链骨架为轴的任何自由旋转，所以 DNA 分子只能通过超螺旋形式释放因解旋而形成的张力。按照惯例，因解旋而引入的超螺旋称为“负超螺旋”。大多数生物体 DNA 都存在负超螺旋，然而，极端嗜热菌中却存在正超螺旋，通过反向 DNA 旋转酶引入正超螺旋，从而使这些生活在沸点温度环境中的生物的 DNA 保持稳定状态。

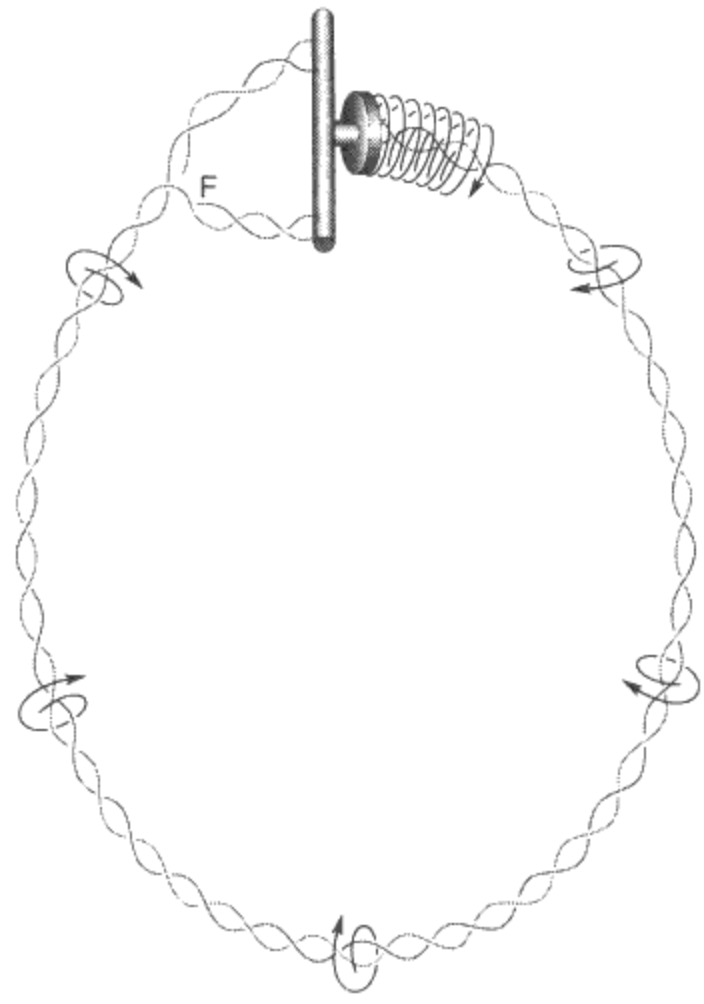


图 20.23 Cairns 旋转概念。当闭合环形 DNA 复制时，两条链必须在复制叉（F）处解旋。因解旋而形成的张力将通过旋转机制得以释放。Cairns 确实将旋转看作是一种主动旋转的机器驱动 DNA 在复制叉处解旋。

可以将 DNA 的超螺旋过程作如下直观化分析：取一条中等偏大的橡皮圈，一只手捏住其顶端，另一只手在橡皮圈的另一端使其向一侧缠绕一圈，你会注意到橡皮圈会对因旋转而引入的张力产生抵抗作用，并通过形成超螺旋（如同“8”字形）而释放张力。旋转的圈数越多，形成的超螺旋数也越多：一个超螺旋数相当于对橡皮圈旋转完整的一圈。向相反方向旋转则会形成相反手型的超螺旋。

如果将橡皮圈被紧握的一端释放，所形成的超螺旋就会获得松弛，在 DNA 分子中，只有将一条链切断，使另一条链能够自由旋转，才能松弛超螺旋。

如果没有其他方式释放张力，那么 DNA 在复制叉处的解旋会引起正超螺旋而不是负超螺旋的形成。因为 DNA 复制时，不需要链的切断而使某一区域的 DNA 发生永久性解旋，迫使其余的 DNA 发生过度缠绕，从而形成补偿性正超螺旋。为直观表述这一过程，观察一下图 20.23 中位于复制叉前方的环形箭头，我们会注意到 DNA 按箭头方式发生缠

绕后，是如何导致箭头后方的 DNA 发生解旋的，而箭头前方的 DNA 则发生过度缠绕。想象一下，将你的手指插入紧邻复制叉的 DNA 链中，然后沿复制移动方向移动你的手指迫使 DNA 的两条链发生分离，你就会想象到链的分离会迫使 DNA 分子按箭头指示方向发生旋转，使 DNA 螺旋过度缠绕。当你的手指沿环形 DNA 链移动时，螺旋产生的张力会对你的手指移动产生越来越强的阻力，所以，因 DNA 在复制叉处解旋而引入的正超螺旋张力必须能够不断地被释放，才不会延缓 DNA 的复制。当你想象在橡皮圈缠绕得越来越紧时是如何抵抗你的缠绕时，你就会理解 DNA 复制时遇到的相同情况。从原理上讲，任何能够释放这种张力的酶都可以发挥旋转作用，事实上，在*E. coli* 细胞所有的拓扑异构酶中，只有 DNA 旋转酶具有旋转功能。

根据断裂 DNA 的单链或是双链而将拓扑异构酶分成两类。第一类拓扑异构酶（I 型拓扑异构酶，如*E. coli* 的拓扑异构酶 I）会使 DNA 单链发生暂时性断裂，而第二类拓扑异构酶（II 型拓扑异构酶，如*E. coli*的 DNA 旋转酶）会在 DNA 分子上产生瞬时的双链断口，随后能够将断口闭合。在 DNA 复制时，为什么*E. coli*的拓扑异构酶 I 不能发挥旋转的功能？原因是该酶只能松弛负超螺旋，不能松弛因复制叉前方因 DNA 复制而引入的正超螺旋。很显然，这些酶在 DNA 链上产生的切口并不能使 DNA 链发生任何方向的自由旋转，但 DNA 旋转酶能够在闭合环形 DNA 分子中引入负超螺旋，从而抵消正超螺旋形成的趋势，因此，DNA 旋转酶发挥旋转体作用。

但也并不是所有的拓扑异构酶 I 都不能松弛正超螺旋，与原核生物类拓扑异构酶 I 不同，来自真核生物和古菌的拓扑异构酶 I（古菌的拓扑异构酶 I 又称类真核生物的拓扑异构酶 I）能够以不同的作用机制，既能松弛负超螺旋，也能松弛正超螺旋。

有直接证据表明 DNA 旋转酶在复制过程具有至关重要的作用。第一，当编码 DNA 旋转酶两条多肽的基因发生突变后，会阻断 DNA 复制，对细胞产生致死效应；第二，新生霉素、香豆霉素、萘啶酸等抗生素能够抑制 DNA 旋转酶的活性，从而阻断 DNA 的复制。

II 型拓扑异构酶的作用机制 Martin Gellert 及同事于 1976 年首次对 DNA 旋转酶进行了纯化。研究者通过测定 DNA 旋转酶在松弛环形 DNA 分子（在本章前面介绍的 colE1 质粒）中引入的超螺旋数来监测其在纯化过程中的活性，然后研究者在反应体系中加入不同量的 DNA 旋转酶及 ATP，1h 后进行凝胶电泳并用能够在 UV 照射下发荧光的 EB（溴化乙锭）进行染色。

图 20.24 显示了该分析实验的结果。一方面，在缺少 DNA 旋转酶（泳道 2）和 ATP（泳道 11）时，基本上只能观察到迁移率较低的松弛环状形式的质粒 DNA 分子的存在；另一方面，当实验者向反应体系中逐渐增加 DNA 旋转酶的添加量后，会观察到迁移率越来越大的具有许多超螺旋的质粒 DNA 分子。在 DNA 旋转酶含量为中等水平的体系中，会产生不同形式的质粒 DNA 分子，对应于凝胶上的不同条带，相应的质粒 DNA 分子具有不同整数的超螺旋。

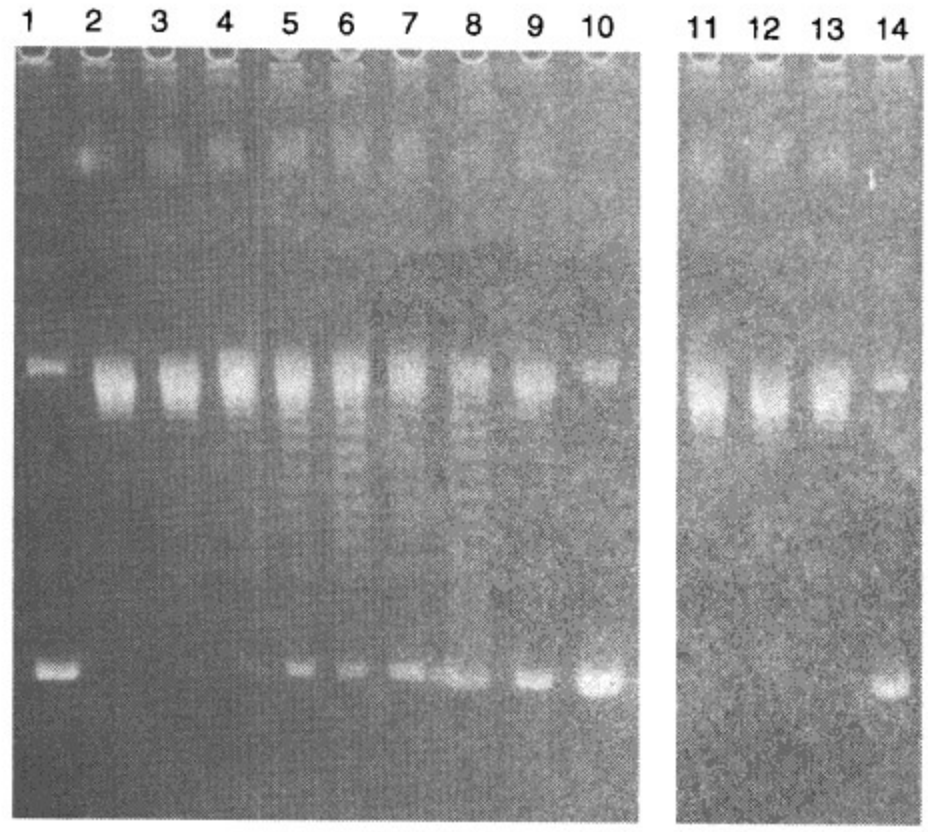
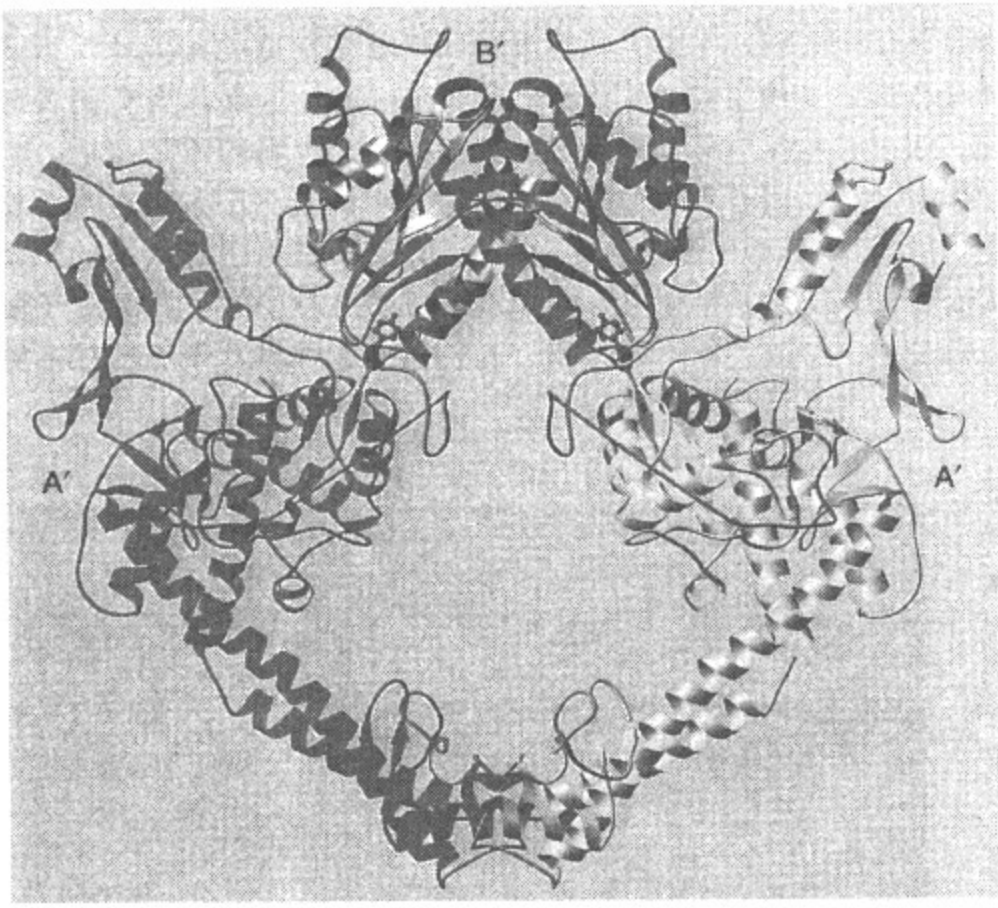


图 20.24 DNA 拓扑异构酶分析。Gellert 及同事将松弛环状 colE1 DNA 与不同数量的 *E. coli* DNA 旋转酶共同温育，同时添加入 ATP、亚精胺和 $MgCl_2$（例外情况已标注）。泳道 1：从*E. coli*细胞中分离的超螺旋 colE1 DNA；泳道 2：无 DNA 旋转酶；泳道 3～10：依次增加 DNA 旋转酶的量，依次为 24ng、48ng、72ng、96ng、120ng、120ng、240ng 和 360ng；泳道 11：无 ATP；泳道 12：无亚精胺；泳道 13：无 $MgCl_2$；泳道 14：在缺少 ATP 的情况下，将超螺旋 colE1 DNA 与 240ng DNA 旋转酶共同温育。（*Source*：Gellert，M.，K. Mizuuchi，M. H. O'Dea and H. A. Nash，DNA gyrase：An enzyme that introduces superhelical turns into DNA. *Proceedings of the National Academy of Sciences USA* 73（1976）fig. 1，p. 3873.）

该实验结果表明 DNA 旋转酶是一种依赖于 ATP 的酶，但所消耗的 ATP 量并不像预期的那样多，磷酸二酯键与再形成。这种适度能量需求的原因在于 DNA 旋转酶自身（不是水分子）是促使 DNA 键发生断裂的试剂，通过形成共价结合的酶-DNA 中间体而使 DNA 键发生断裂。该中间体保存了 DNA 磷酸二酯键中的能量，可被重新用于 DNA 末端的再连接，同时 DNA 旋转酶以原有的形式被释放。

有什么证据能够证明酶-DNA 中间体的存在呢？James Wang 及同事在 DNA 断裂-重新连接的循环过程的中间环节时将酶进行变性处理，收集到了 DNA-酶复合体。在该复合体中，DNA 的两条链均存在切口，两切口相距 4 个碱基对，呈交错结构，DNA 旋转酶与这两个突出的 DNA 末端发生共价结合。1980 年，Wang 及同事通过 DNA 旋转酶上的一个酪氨酸残基进一步证明了酶与 DNA 间的共价结合，研究者将 [^{32}P] DNA 与 DNA 旋转酶共同温育，然后通过使酶变性从而捕获 DNA-旋转酶复合体，分离纯化出该复合体后，用核酸酶将复合体中的 DNA 消化殆尽，分离得到了标记位于 A 亚基上的 [^{32}P] 酶（DNA 旋转酶与所有的原核生物 DNA 拓扑异构酶 II 一样，是由两个不同亚基构成的四聚体结构：A_2B_2）。旋转酶的亚基 A 被 ^{32}P 标记的事实更有力地表明该亚基通过一个氨基酸残基而与 ^{32}P [DNA] 相连。那么，哪一个氨基酸残基能够与 DNA 发生连接？Wang 及同事将标记的 A 亚基于煮沸的 HCl 中进行消化，使之降解为游离的氨基酸组分，然后纯化被标记的氨基酸，与之共纯化的是磷酸酪氨酸，所以，通过 A 亚基上的酪氨酸残基，旋转酶与 DNA 实现了共价结合。

DNA 旋转酶以及其他 DNA 拓扑异构酶 II 是如何向 DNA 分子中引入负超螺旋的？最简单的解释是：这些酶能够使双螺旋的一部分穿越另一部分。图 20.25 显示的是酵母拓扑异构酶 II 的 X 射线晶体结构，和所有形式的真核生物拓扑异构酶 II 一样，是同源二聚体结构，每一单体都包含有相当于原核生物拓扑异构酶 II 的 A 和 B 亚基的结构域。酵母拓扑异构酶 II 是由 2 个新月形单体构成的心形蛋白质，呈双颚结构，分别位于上部与底部。



二聚体基本界面

图 20.25　酵母拓扑异构酶 II 的晶体结构。左侧的单体分别用绿色和橙色表示，右侧的单体分别用黄色和蓝色表示。在绿色和黄色单体上存在着相当于原核生物拓扑异构酶 A 亚基的功能域（标注为 A′），在橙色和蓝色单体上存在着相当于原核生物拓扑异构酶 B 亚基的功能域（标注为 B′）。具有 ATP 酶活性的 B′功能域形成酶的上颚，A′功能域形成酶的下颚。图中显示的结构中，酶的双颚是关闭的。在反应过程中与 DNA 发生连接的酪氨酸活性位点用紫色六边形表示，位于 A′结构域和 B′结构域交界处附近，在底部标注了单体间的主要接触面。（*Source*: Wang, J. C., Moving one DNA double helix through another by a type II DNA topoisomerase: the story of a simple molecular machine. *Quarterly Reviews of Biophysics* 31, 2 (1998), f. 8, pp. 107-144. © 1998 Cambridge University Press.）

图 20.26 展示了在 DNA 片段相互穿越过程中双颚协同作用的模型。上颚与 DNA 的 G 片段结合，该片段包含有其他片段穿越的入口，被 ATP 激活后，上颚又与 DNA 的 T 片段结合，该片段能够穿越 G 片段而被转运。G 片段与 T 片段相互垂直，DNA 旋转酶断裂 G 片段使之产生入口，当 T 片段穿越 G 片段进入下颚后，被释放出来。

小结　存在一种或多种被称为解旋酶的酶，可利用 ATP 能使两条亲本 DNA 链在复制叉处解旋分离。当解旋酶对闭合环形 DNA 的两条亲本链解旋时，会在 DNA 分子中引入抵消正超螺旋的张力，必须克服这种张力，否则复制叉的行进将会受阻。旋转可以释放这种张力，在 *E. coli* 中，DNA 旋转酶是具有旋转功能的首选者，通过在复

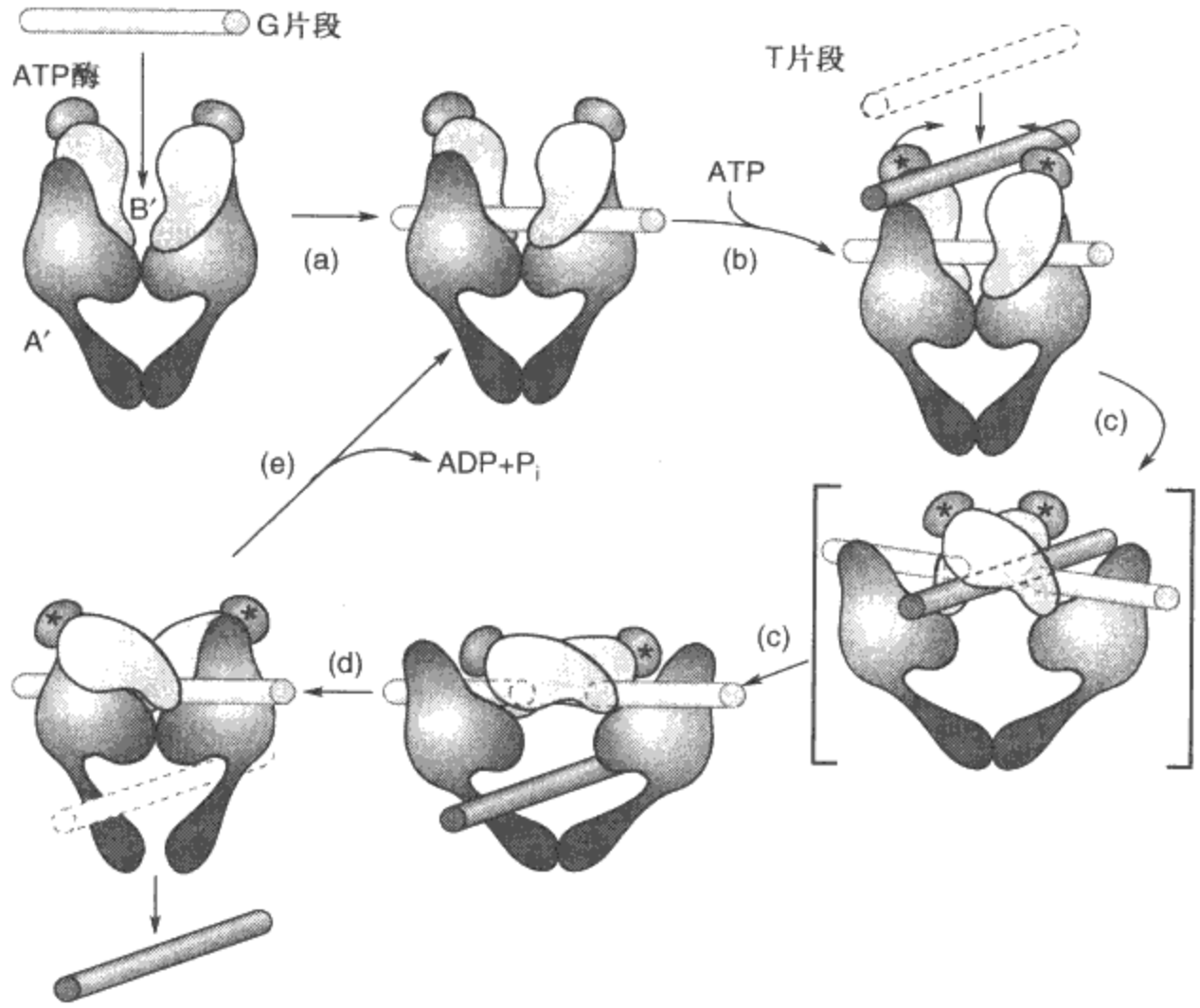


图 20.26　拓扑异构酶 II 催化的反应中，DNA 片段穿过步骤模型。基于拓扑异构酶 II 的晶体结构以及其他方面的证据，Wang 及同事提出了如下模式：(a) 酶的上颚打开，与 DNA 的 G 片段（双链 DNA）结合，该片段能够断裂形成允许其他 DNA 片段穿过的门。DNA 结合后导致酶的构象发生改变，引导 B′功能域的酪氨酸活性位点进入攻击 DNA 的位置。(b) 每个上颚中的 ATP 酶功能域与 ATP（用星号表示）结合，此外，上颚还与 DNA 的双链 T 片段结合，T 片段将会穿过 G 片段。(c) 通过包括假定中间体（括号内）在内的一系列构象变化，活性位点将 G 片段切断，使 T 片段穿过 G 片段并进入到下颚中。在 c 步骤中，B′功能域的前面部分是透明的，以便可以观察其后的 DNA。(d) 下颚张开，释放 T 片段，而 G 片段被重新接合。(e) 拓扑异构酶 II 水解 ATP，使其恢复到能够接受另一个 T 片段的状态，并重复这一片段穿越过程。(*Source*: Adapted from Berger, J. M., S. J. Gamblin, S. C. Harrison and J. C. Wang, Structure and mechanism of DNA topoisomerase II. *Nature* 379: 231, 1996.)

制的 DNA 中引入负超螺旋而中和正超螺旋，否则正超螺旋会阻断 DNA 的复制。

20.3　DNA 损伤与修复

生物体中有很多不同的情况可导致 DNA 损伤，如果损伤的 DNA 不能得到及时的修复，就会引起突变，即 DNA 碱基序列的改变。首先值得强调一点的是 DNA 损伤与 DNA 突变是两个不同的概念。DNA 损伤仅是 DNA 简单的化学变化，而 DNA 突变则是 DNA 碱基对的改变。例如，由 G-C 碱基对改变为甲基化 G-C 碱基对属于 DNA 损伤范畴；而由 G-C 碱基对转变为其他正常的碱基对（如 A-T、T-A 或 C-G）则属于 DNA 突变。事实上，在 DNA 复制过程中，甲基化 G 不会与 C 配对，而可能会与 T 发生错配，如果情况确实如此，那么在进行第二轮 DNA 复制时，甲基化的 G 便会取代 A 而与 T 发生错配，从而将正常的 G-C 碱基对转变为 A-T 碱基对，导致突变的发生。这个事例说明，在 DNA 损伤向突变转化的过程中，DNA 复制起到了重要作用。

由烷基化试剂引起的碱基修饰以及由紫外线辐射引起的嘧啶二聚体就是两种最普遍的 DNA 损伤事例。下面我们将具体学习原核生物与真核生物细胞修复 DNA 损伤的机制，在多数 DNA 损伤修复机制中均涉及 DNA 的复制。

碱基的烷基化修饰引起的 DNA 损伤

在自然界中存在一些天然的或人工合成的亲电试剂，能够进攻其他分子的负电中心并与之结合。有许多物质在体内通过代谢作用后成为亲电试剂，DNA 是生物体内明显带有负电中心的物质，每个核苷酸分子包含带有一个负电荷的磷酸基团和带有部分负电荷的碱基，当亲电试剂进攻这些负电中心时，通常会将含碳的基团，即烷基添加到负电中心，这一过程称之为烷基化（alkylation）。

图 20.27 描述了 DNA 分子负电中心的特征。除了磷酸二酯键、鸟嘌呤的 N7 和腺嘌呤的 N3 是亲电试剂易于进攻的位点外，还有许多其他易于受到亲电试剂进攻的位点，而且不同的烷基化试剂有各自不同的进攻靶位点。

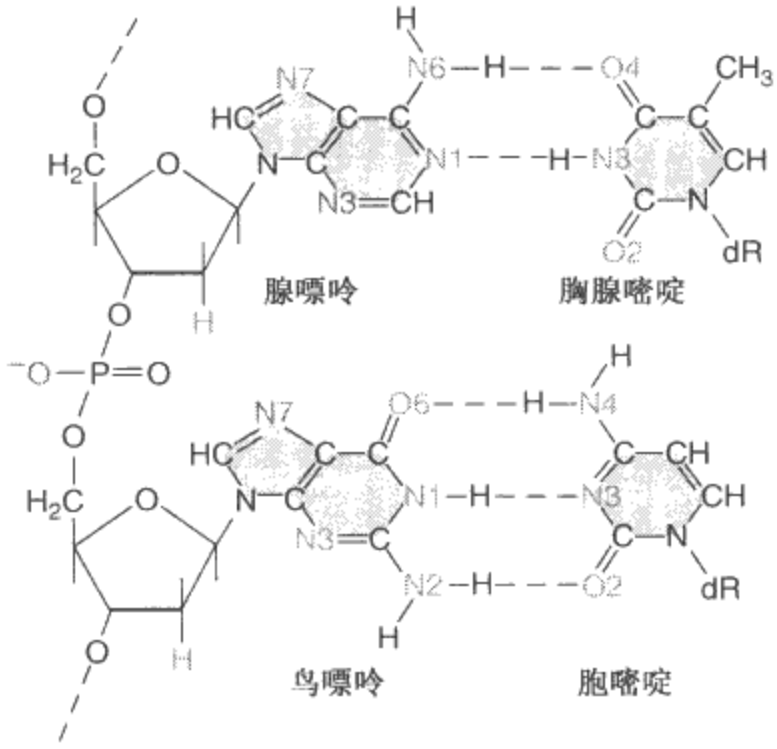


图 20.27　DNA 分子内的富集电子中心。亲电子试剂最常攻击的位点是磷酸基团、鸟嘌呤的 N7 和腺嘌呤的 N3（红色），其他靶位点用蓝色表示。

当 DNA 分子上的这些位点发生烷基化修饰后会出现怎样的后果？首先看一下鸟嘌呤的 N7 和腺嘌呤的 N3 这两个主要位点发生烷基化修饰的后果。鸟嘌呤的 N7 位点发生烷基化后不会改变其与胞嘧啶发生碱基互补配对的性质，一般是无害的；而腺嘌呤的 N3 位点发生烷基化后，其后果比较严重，因为烷基化修饰后产生的非编码碱基（noncoding base）3-甲基腺嘌呤（3mA）不能与其他碱基进行正确配对。由于 DNA 聚合酶不能准确识别含有 3mA 的碱基对，DNA 聚合酶就会在 3mA 损伤处停下来而阻碍了 DNA 的复制，阻断 DNA 复制对细胞而言是致死的，所以认为 3mA 对细胞是有毒性的。另一方面，也正是我们在本章后面内容中将要学习到的，这种 DNA 复制的暂停可以在损伤不被修复的情况下解除，这种 DNA 复制恢复机制是有错误倾向的，因此有可能会引起突变。

此外，所有参与碱基互补配对的氮原子和氧原子也会发生烷基化修饰（图 20.27），会直接破坏碱基配对而导致突变。鸟嘌呤的 O6 位点发生烷基化修饰后最可能会导致突变，尽管该位点很少受到亲电试剂的进攻。该位点一旦发生烷基化修饰就会产生很强的诱变性，因为烷基化的鸟嘌呤会与胸腺嘧啶而不是胞嘧啶互补配对。例如，常用的诱变剂乙基甲烷磺酸（EMS）能够将乙基基团（CH_3CH_2）转移到 DNA 分子上，从而使鸟嘌呤的 O6 位点发生烷基化（图 20.28）。鸟嘌呤 O6 位点的烷基化修饰改变了鸟嘌呤的互变异构体形式（双键形式），而使其与胸腺嘧啶进行碱基互补配对，导致 G-C 碱基对转换为 A-T 碱基对。

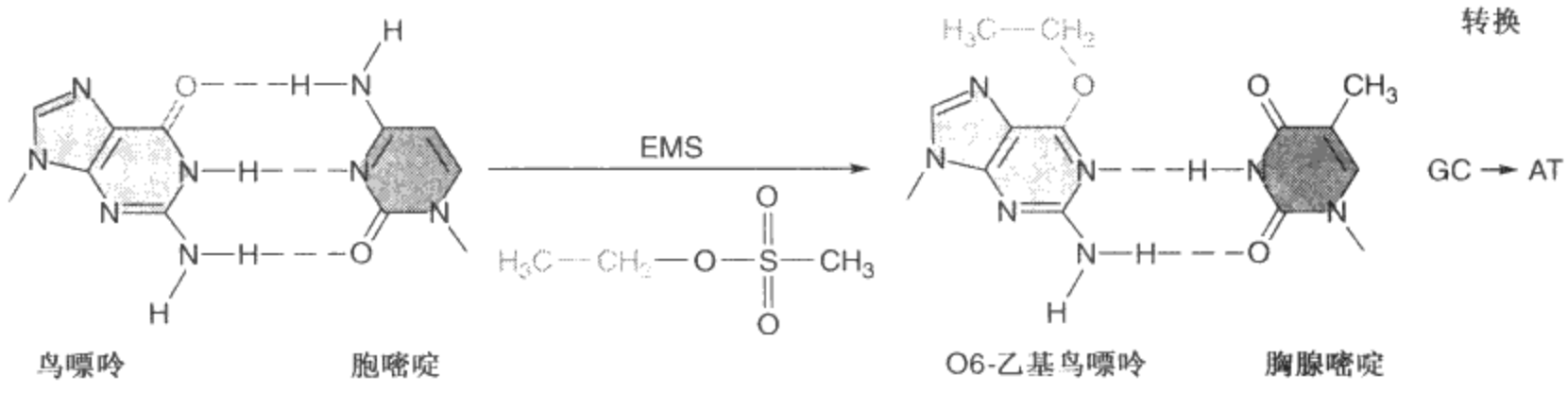


图 20.28　EMS 对鸟嘌呤的烷基化修饰。左边是正常的鸟嘌呤-胞嘧啶碱基对，值得注意的是，鸟嘌呤的自由 O6 原子（红色），EMS 作为供体将乙基转移至鸟嘌呤的 O6 原子上，形成的 O6-乙烷基鸟嘌呤（右），能够与胸腺嘧啶而不是胞嘧啶进行互补配对。经过又一轮的 DNA 复制后，A-T 碱基对取代了原来的 G-C 碱基对。

环境中有许多致癌物质（carcinogen）都是亲电试剂，能够使 DNA 发生烷基化修饰。正如我们刚才所了解到的那样，这会导致突变的发生，如果控制或影响细胞分裂的基因发生突变，就会使细胞失去控制复制的能力而转变为癌细胞。

小结　乙基甲烷磺酸（EMS）等烷化剂能够使 DNA 碱基发生烷基化修饰，不改变碱基互补配对性质的烷基化是无害的，而引起 DNA 复制停止的烷基化则是有害的，如果损伤的 DNA 没有得到修复而细胞仍继续复制就会导致突变，有些烷基化会改变碱基互补配对的性质，所以是诱变剂。

紫外线辐射引起的 DNA 损伤

紫外线（ultraviolet，UV）辐射会引起同一条 DNA 链上相邻的嘧啶碱基发生交联而形成嘧啶二聚体（pyrimidine dimmer）。图 20.29 示意了胸腺嘧啶二聚体的结构，并阐明了胸腺嘧啶二聚体是如何破坏 DNA 双链间的碱基互补配对的。这种二聚体之所以能够阻断 DNA 复制，是因为复制机构不能确定将哪一种碱基引入到与二聚体相对应的位置处。正如我们将要学习到的那样，有时 DNA 的复制过程会继续进行下去，引入的碱基不会准确地进行碱基互补配对，如果插入的碱基是错误的则会引起突变。

紫外辐射具有重要的生物学意义。太阳光中含有紫外线光谱，大多数生命有机体都在某种程度上暴露于紫外光照射之下。紫外光辐射具有的诱变作用也就解释了为什么日光能够引发皮肤癌：日光中的紫外成分损坏皮肤细胞中的 DNA，从而导致突变的发生，致使这些细胞无法调控自身的分裂。虽然紫外线具有一定的危险性，但我们很幸运，地球上方大气层中的臭氧层作为保护罩，能够吸收大量的紫外线。然而，科学家注意到，臭氧层已经出现了漏洞，其中最大的漏洞出现在南极洲上方。关于引起臭氧消耗的原因还存在着争论，但这些原因中包括在空调及塑料使用过程中向大气层中释放的化合物。如果我们不停止破坏臭氧层的行为，人类将遭受更多的包括皮肤癌在内的紫外线损伤。

γ 射线及 X 射线引起的 DNA 损伤

同紫外线一样，高能 γ 射线及 X 射线也能与 DNA 分子发生直接的相互作用，但它们对 DNA 的损伤主要是由于产生的电离分子（如水）包围在 DNA 分子周围。自由基就是带有不成对电子的化学物质，这些自由基特别是含氧的自由基具有很高的活性，能够快速与相邻的分子发生作用。当自由基与 DNA

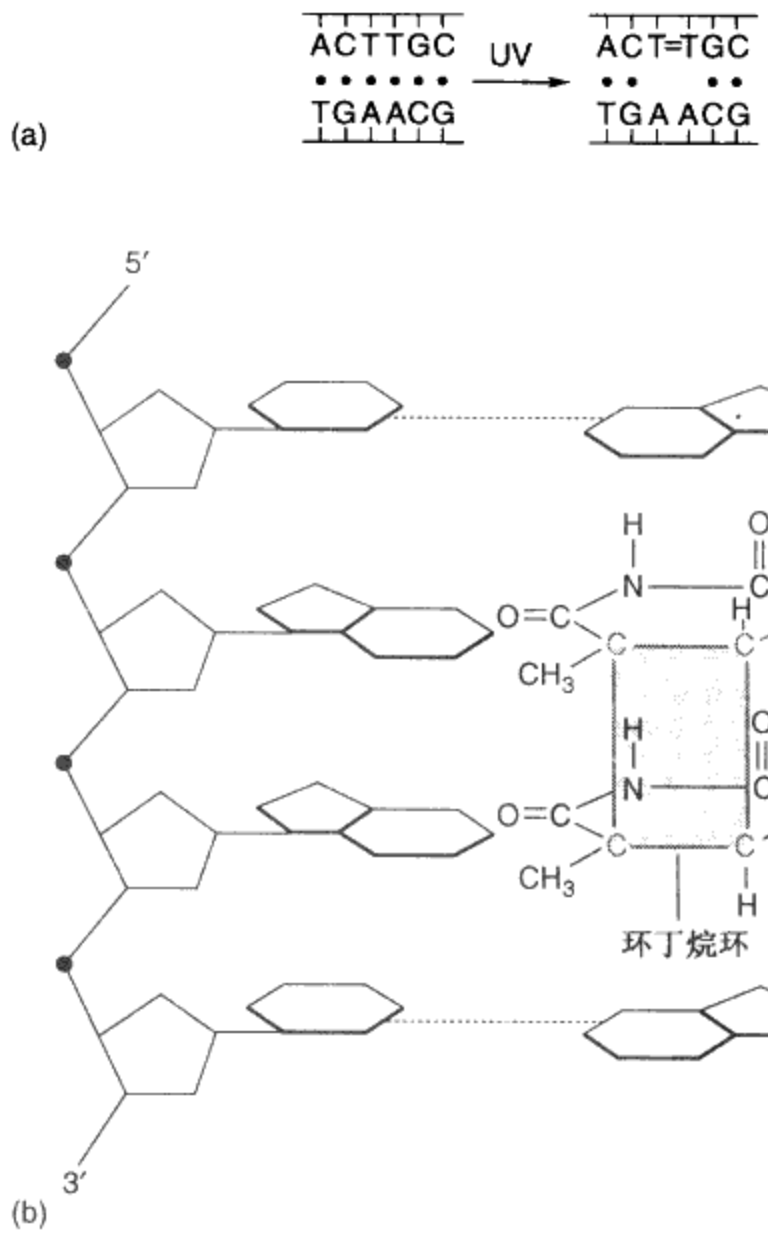


图 20.29　嘧啶二聚体。（a）紫外线照射使上链中相邻的两个嘧啶交联在一起（图中所示例的为胸腺嘧啶），这将导致 DNA 扭曲变形，使这两个非编码碱基不再与腺嘌呤配对；（b）通过两个化学键将相邻的两个嘧啶连接在一起，形成环四丁烷结构（红色）。

分子发生作用时，可以引起碱基的改变，但更多的情况是引起 DNA 单链或双链的断裂。

DNA 碱基至少会遭受到 20 种氧化损伤，特别是因电离辐射或正常氧化代谢而产生的活性氧。对 8-氧代鸟嘌呤（oxoG）即 8-羟基鸟嘌呤（图 20.30）这种氧化碱基研究得最深入，原核生物或真核生物的 DNA 聚合酶会将 oxoG 误读为胸腺嘧啶，从而在本应插入胞嘧啶的位置处插入腺嘌呤，形成 oxo G-A 碱基对。这两个碱基都具有诱变效应，如果这两个碱基在下一轮 DNA 复制前不能被移除，就可能会引发突变。单链断裂的后果并不严重，因为很容易修复，只要把断开的链末端连接起来即可，而 DNA 双链断裂后就很难准确修复，从而会引发永久性突变。因为电离辐射能够使染色体断裂，所以电离辐射即是一种诱变剂，又是一种断裂剂（clastogen）。

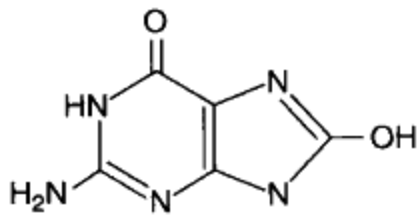


图 20.30　8-氧代鸟嘌呤（8-羟基鸟嘌呤）。

小结　不同种类的辐射会造成不同的 DNA 损伤。紫外线的能量相对较低，能够产生嘧啶二聚体等中等程度的 DNA 损伤；而 γ 射线及 X 射线为高能射线，能够使 DNA 周围的分子呈离子化状态，产生能够进攻 DNA 的高活性自由基，引起 DNA 碱基的改变或链的断裂。

DNA 损伤的直接修复

处理 DNA 损伤的方法之一就是进行修复，将损伤恢复到原来的正常状态。有两条基本修复途径：①直接去除损伤；②移去损伤的 DNA 片段，然后补上新的未受损伤的 DNA。现在我们将具体了解一下 *E. coli* 细胞直接修复 DNA 损伤的两种方法。

在 20 世纪 40 年代后期，Albert Kelner 以链霉菌为研究对象，试图阐明温度对 DNA 紫外损伤修复的影响，然而，他观察到在相同的温度条件下，有些细菌孢子对损伤修复的很快。很明显，除了温度条件外，还有其他因子在起作用。最后，Kelner 注意到损伤修复比较快的细菌孢子是由于暴露在实验室的光照之下，当控制实验室条件，把细菌孢子培养在黑暗条件下时，不会再有损伤修复事件的发生。不久，Renato Dulbecco 以侵染含紫外损伤的噬菌体的细菌为研究对象进行实验时，观察到了相同的现象。这就是大多数生命有机体均具有的一种重要的损伤修复机制——光复活（photoreactivation）或光修复（light repair），然而，包括人类在内的胎盘哺乳动物却没有光复活途径。

在 20 世纪 50 年代后期，人们发现了催化光复活反应的酶——光复活酶（photoreactivating enzyme）或光裂合酶（DNA photolyase），图 20.31 描述了该酶催化光复活反应的机制。首先，光复活酶识别 DNA 损伤位点并与之结合（嘧啶二聚体），然后酶吸收 UV-A 到蓝光光谱范围内的光后被激活，从而断裂嘧啶二聚体之间的化学键，使嘧啶碱基恢复到原来单体状态，最后酶与 DNA 解离，损伤得到修复。

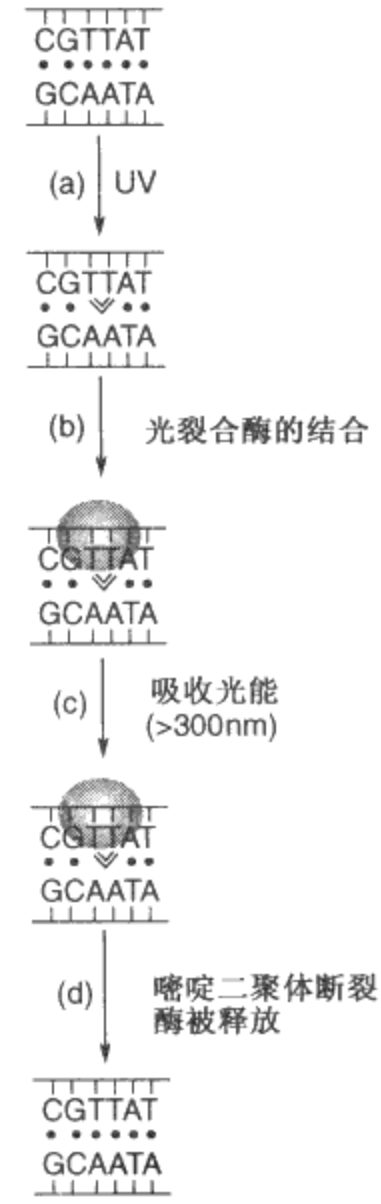


图 20.31　光复活作用模型。（a）紫外线照射引发嘧啶二聚体的形成；（b）DNA 光裂合酶（红色）与 DNA 的损伤区域结合；（c）酶吸收近紫外到可见光光谱范围内的光；（d）酶将嘧啶二聚体断裂，然后与已被修复的 DNA 解离。

从 *E. coli* 到人类，所有的生命有机体都能够直接逆转另一种损伤，即鸟嘌呤 O6 位点的烷基化损伤。当 DNA 被甲基化或乙基化修饰后，存在于生物体内的 O6-甲基鸟嘌呤甲基转移酶（O6-methylguanine methyl transferase）能够催化这种损伤修复

反应，它能直接将甲基或乙基从鸟嘌呤的 O6 原子上转移到酶蛋白分子的一个半胱氨酸残基上，从而修复受损伤的 DNA，修复机制如图 20.32 所示。

烷基的接受位点是酶蛋白质的一个半胱氨酸残基上的硫原子，严格来讲，这意味着甲基转移酶不是真正意义上的酶，正常的化学酶在其所催化的反应完成后酶分子不会发生任何改变。而 O6-甲基鸟嘌呤甲基转移酶的活性在反应完成后不能恢复，所以我们把这类酶称之为是一种自杀性酶，因为其在履行完自己的功能后便失活。因此，这种修复过程是需要付出代价的，每发生一次修复事件就需要消耗一个蛋白质分子。

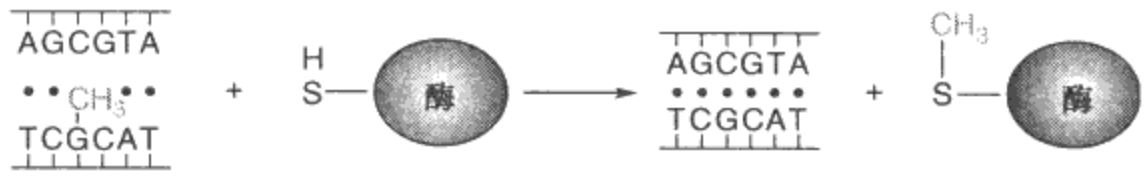


图 20.32　O6-甲基鸟嘌呤甲基转移酶的作用机制。酶的巯基接受 DNA 中鸟嘌呤的甲基（蓝色），使酶自身失活。

O6-甲基鸟嘌呤转移酶的另一个特性也很值得注意，该酶能够被烷基化的 DNA 所诱导，至少在 *E. coli* 细胞中是这样的，这表明一直处于烷化剂环境中的细菌细胞比刚处于诱变剂环境中的细菌细胞对 DNA 损伤具有较大的抗性。

小结　因紫外线辐射造成的 DNA 损伤（嘧啶二聚体）可以直接被 DNA 光裂合酶所修复，此酶利用近紫外至蓝光波段的光而将嘧啶二聚体中化学键打开；鸟嘌呤 O6 位点的烷基化可以被自杀性 O6-甲基鸟嘌呤甲基转移酶修复，可以将烷基基团转移到酶自身的一个氨基酸残基上。

切除修复

通过直接逆转修复方式只能修复少量的 DNA 损伤，因为大多数损伤既不是嘧啶二聚体损伤，也不是 O6-烷基鸟嘌呤损伤，所以必须通过不同的机制来修复这些损伤。切除修复（excision repair）能够消除大多的 DNA 损伤，首先将受损伤的 DNA 切除，然后用新合成的 DNA 进行替换。切除修复机制包括碱基切除修复和核苷酸切除修复。碱基切除修复机制较为普遍，主要针对碱基改变轻微的 DNA 损伤，如细胞试剂引起的化学修饰；核苷酸切除修复机制主要修复碱基发生重大变化的 DNA 损伤，其中许多损伤会引起 DNA 双螺旋结构的扭曲，这些改变主要是由于细胞外的诱变剂造成的，UV 辐射引起的嘧啶二聚体就是这类损伤的一个典型事例。

碱基切除修复　在碱基切除修复（base excision repair，BER）中，由 DNA 糖基化酶（DNA glycosylase）识别受损碱基，通过使 DNA 发生扭曲而使受损碱基从碱基对中凸出来，然后水解受损碱基与核糖之间的糖苷键（glycosidic bond）（图 20.33），从而除去受损碱基，产生无嘌呤（apurinic）或无嘧啶（apyrimidinic）的 AP 位点，即没有嘌呤碱基或嘧啶碱基的核糖位点。AP 内切核酸酶（AP endonuclease）能够识别 AP 位点，在 AP 位点的 5′端一侧将 DNA 链切断或造成缺口（内切核酸酶中“endo”是指该酶能够在内部而不是在游离的末端将 DNA 链切断）。在 *E. coli* 中，DNA 磷酸二酯酶在 AP 位点处将磷酸戊糖切除，然后由 DNA 聚合酶 I 进行修复合成，按 5′→3′方向，在降解 DNA 的同时合成新的 DNA 片段，但 DNA 聚合酶不能修复缺口，最后由 DNA 连接酶将切口连接起来，完成损伤修复。不同的损伤碱基可由不同的 DNA 糖基化酶识别，人类至少有 8 种 DNA 糖基化酶。由于 DNA 复制一般不会因碱基受到微小的化学修饰而停止，但可能会引起错码，所以在防止发生突变方面，BER 发挥着重要作用。

真核生物的大多数 BER 过程（图 20.34a～e）与原核生物相似，只是没有 DNA 磷酸二酯酶的参与。AP 位点被切开后，由 DNA 聚合酶 β 填补空隙，同时将磷酸戊糖切除（蓝色），但该修复方案存在一个基本问题：细菌的 DNA 聚合酶 I 具有编辑校正活性，而 DNA 聚合酶 β 则无此功能，是一种具有错误倾向的酶，平均复制 4000 个核苷酸就会出现 1 次错误，而且自身又不能修复错误。这种情况听起来可能不是很糟糕，但考虑到人类的基因组中每天可能有 2000～8000 个碱基会受到损伤，如果按 DNA 聚合酶 β 的这种错误概率计算的话，BER 系统每天就会在人类的基因组中引入 5～20 次突变。

幸运的是，真核细胞具有解决这个问题的方案。2002 年，Kai-Ming Chou 和 Yung-Chi Cheng 研究表明，人类的无嘌呤/无嘧啶（AP）内切核酸酶（APE1）能够与 DNA 聚合酶 β 共同作用，以校正 DNA 聚合酶 β 产生的错误。已知酵母的 APE1 除了具有主要的内切核酸酶活性外，还有 3′→5′外切核酸酶活性，但 3′→5′外切核酸酶活性太弱可能没

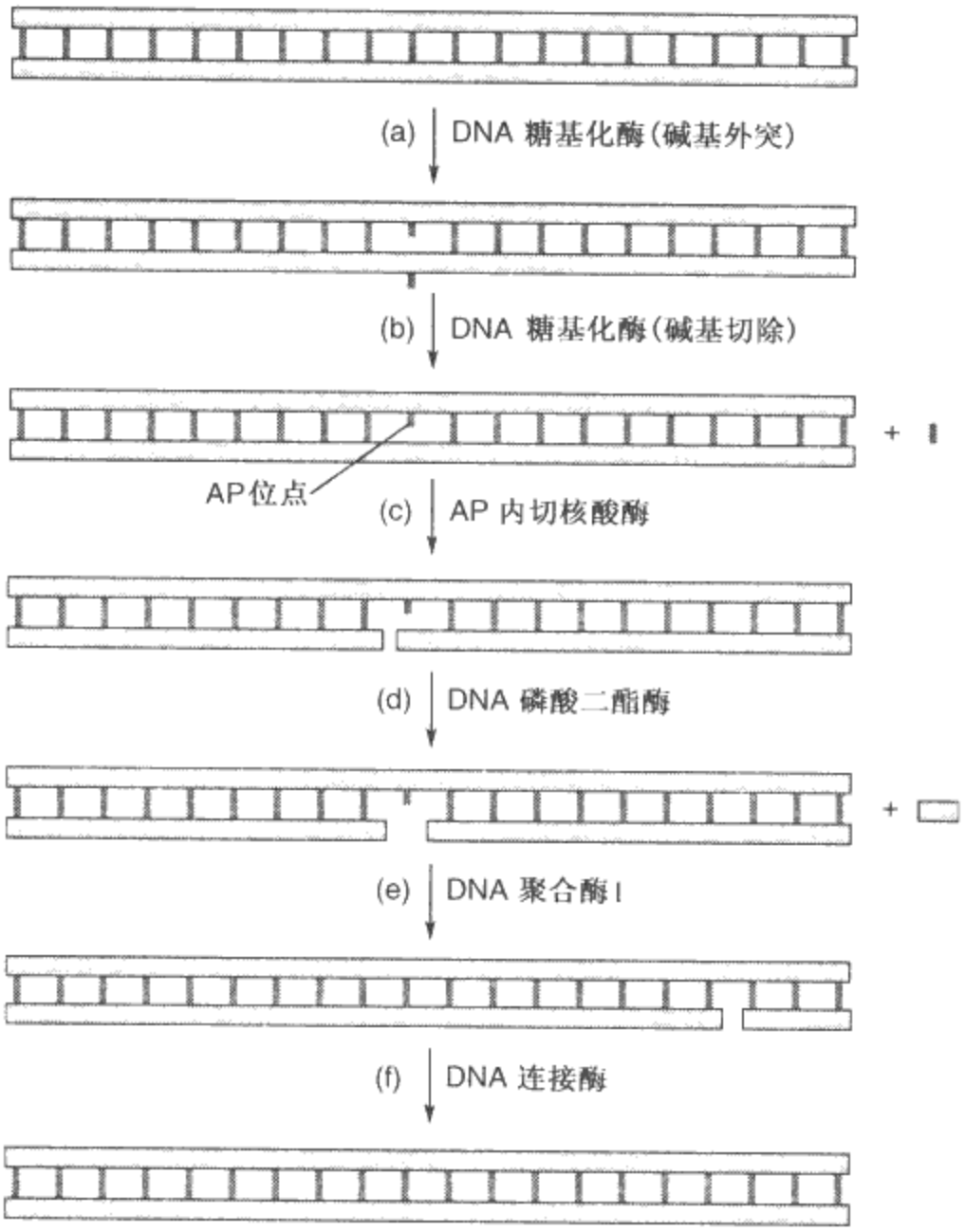


图 20.33　*E. coli* 的碱基切除修复。(a) DNA 糖基化酶使受损碱基（红色）外突；(b) DNA 糖基化酶切除受损的外突碱基，在下方 DNA 链伤留下一个无嘌呤或无嘧啶位点；(c) AP内切核酸酶从 AP 位点 5′端一侧切断 DNA 链；(d) DNA 磷酸二酯酶去除 DNA 糖基化酶留下的 AP-脱氧核糖磷酸（右边的黄条块）；(e) DNA聚合酶 I 将填补空隙，并继续向下游修复合成几个核苷酸，DNA 降解的同时将其修复；(f) DNA 连接酶黏合 DNA 聚合酶 I 留下的缺口。

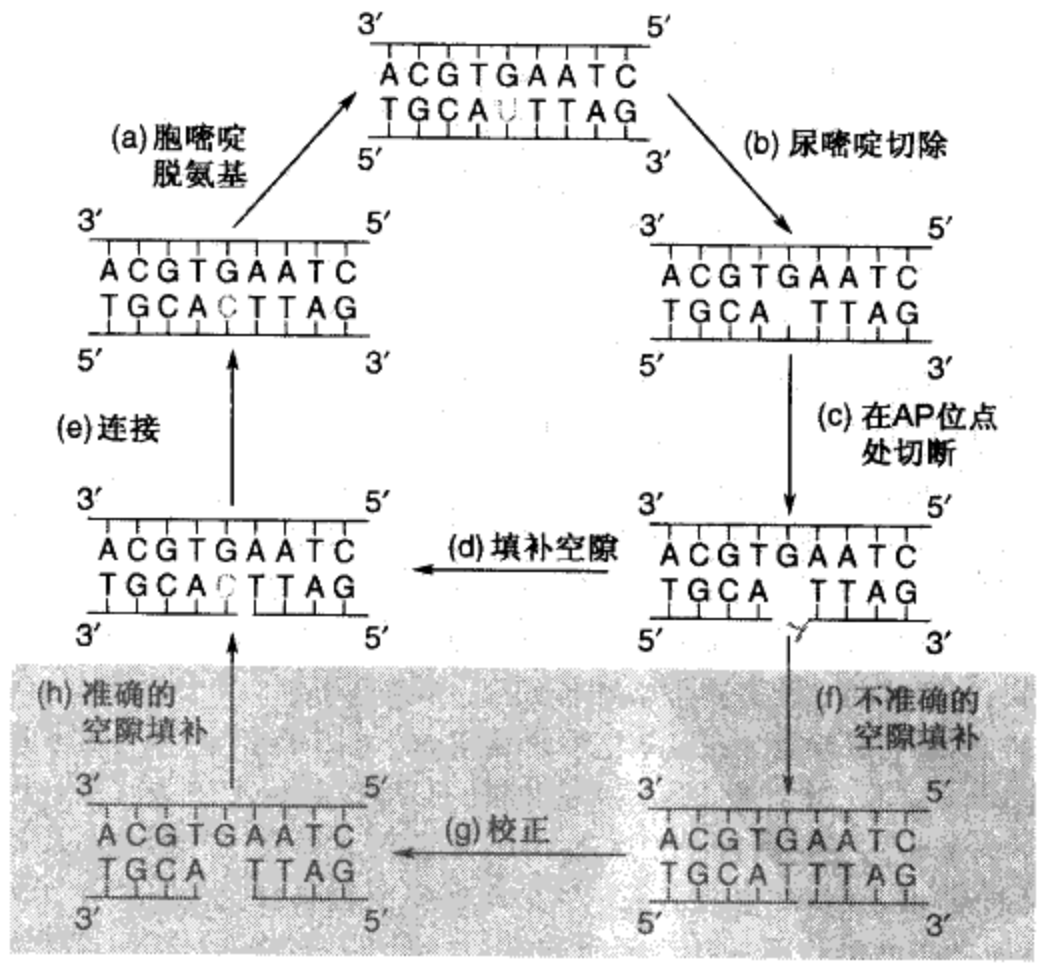


图 20.34　人类的 BER 途径。(a) DNA 分子的下方链中，胞嘧啶发生自发脱氨基作用，使 C（蓝色）转变为 U（橙黄色）；(b) 糖基化酶将尿嘧啶切除；(c) APE1 在无嘧啶位点的 5′端一侧将 DNA 链切断；(d) DNA 聚合酶 β 用 C（蓝色）准确地填补空隙，同时将核糖磷酸切除（绿色）；(e) DNA连接酶 I 黏合切口，使 DNA 恢复到正常状态；(f) DNA 聚合酶偶尔也会出现错误，可能会掺入错误的 T（红色），而不是正确的 C，在错配碱基的 3′端留下一个切口；(g) APE1 利用其 3′外切核酸酶活性切除错配的 T，产生空隙；(h) DNA 聚合酶 β 准确地掺入能够与 G 配对的 C（蓝色），修复错配，然后将 DNA 连接起来，即可恢复到正常状态。

有什么生物学意义。Chou 和 Cheng 的研究表明，对准确配对的核苷酸而言，酵母 APE1 的 3′→5′外切核酸酶活性确实很弱，但遇到末端错配的核苷酸，如当 DNA 聚合酶 β 没有准确填补空隙时，其3′→5′外切核酸酶活性会提高 50～150 倍（图 20.34f）。

当相邻的两条 DNA 链中的一条其末端拥有错配碱基时，DNA 连接酶 I 不能有效地将这两条 DNA 链连接在一起，如图 20.34 步骤 f 中所示的结构。事实上，DNA 连接酶 I 对末端错配 DNA 链的连接效率低于 10%。如果 APE1 确实能够修复由 DNA 聚合酶 β 产生的错配 DNA，那么可以预测：APE1 能够与连接酶协同作用，修复错配的碱基并有效地将 DNA 链连接起来。Chou 和 Cheng 利用由纯化的 DNA 连接酶 I、DNA 聚合酶 β 和 APE1 构成的重组系统，证明 APE1 能够有效地提高连接效率，依据其浓度，连接效率提高 10%～95%。所以，APE1 确实能够修复由 DNA 聚合酶 β 产生的错配。

碱基切除修复的一个特殊事例是对 8-氧代鸟嘌呤的修复，在本章前面的内容中我们已了解到，oxoG 是 DNA 受到氧化损伤的结果。oxoG 可与 A 配对，形成 oxoG-A 碱基对，这两个碱基都具有诱变效应，在下一轮 DNA 复制过程中都可能发生错配而引发突变。在人类中，这些突变会导致癌症的发生。但是有氧生物进化获得了处理这些错配碱基的机制。

Gregory Verdine 及同事于 2004 年阐释了错配碱基 A 的修复机制，参与修复的酶为腺嘌呤 DNA 糖基化酶（adenine DNA glycosylase），包括细菌中的 MutY 和人类的 hMYH。该酶能够切除与 oxoG 发生错配的 A，但不会切除正确配对的 C，而且也不会对所有与 T 正确配对的 A 产生作用。腺嘌呤 DNA 糖基化酶如何区分这些碱基的？对 MutY 与包含 oxoG 的模式 DNA 分子组成的复合体进行 X 射线晶体学分析可能会找到问题的答案，但这些复合体很不稳定，难以获得其结晶体。因此，Verdine 及同事使包含 oxoG 的寡核苷酸分子与 MutY 通过二硫键共价结合在一起，从而形成紧密结合的复合体，且易形成结晶体。

复合体的晶体结构显示，oxoG-A 碱基对与酶之间会发生紧密而特异的相互作用，腺嘌呤碱基失去与 oxoG 间的相互作用，外突进入酶的活性位点。腺嘌呤与脱氧核糖间的糖苷键断裂，导致腺嘌呤从 DNA 上解离。相反，正常的 T-A 碱基对不会与酶发生紧密而特异的相互作用，所以这些正常的碱基对不会受到酶的作用。而且，oxoG-C 碱基对也会像 oxoG-A 碱基对一样，与酶同样的相互作用，但 C 不会发生外突，不能进入酶的活性位点，因此 C 也不会被切除。

oxoG 是如何被切除的呢？其 BER 过程由 DNA 糖基化酶 oxoG 修复酶（oxoG repair enzyme）起始，该酶能够切断 oxoG 与脱氧核糖之间的糖苷键。在人类中，oxoG 修复酶为 hOGG1，能够识别并区分 oxoG-C 碱基对与 G-C 碱基对，使 oxoG 外突，然后将其切除。关键的问题是，hOGG1 如何从众多的正常 G-C 碱基对中区分出稀少的 oxoG-C 碱基对？

实际上，oxoG-C 碱基对与 G-C 碱基对均可作为 hOGG1 的底物，使 G 和 oxoG 均从 DNA 双螺旋中外突出来，但 G 和 oxoG 进入酶的不同位点，只有 oxoG 能够进入糖基化酶的活性位点，然后被切除。但这个结论又很难定论，因为必须捕获到正常的 G 发生外突的情况，而且这种外突一般都是十分短暂的。Verdine 及同事使用与捕获 MutY-oxoG-A 复合体相同的二硫键交联策略，捕获外突的 G。

图 20.35 显示了包含有 oxoG-C 碱基对及正常 G-C 碱基对复合体的晶体结构，在这 2 个复合体中，远离的 C（失去外突的配对者）与蛋白质的同一位点发生紧密结合。很显然，糖基化酶能够识别 C，引起酶的暂停，使与 C 配对的碱基外突以判断该碱基是否正常。oxoG 与 G 都会从 DNA 双螺旋中突出出来，但 oxoG 的结合位点（可能是损伤识别位点）与 G 的结合位点截然不同。Verdine 及同事认为，在 oxoG 外突进入损伤识别位点的过程中，G 复合体是系列结构转换的最后中间体，由于化学构象存在细微差别，所以 G 不会经历转换的最后步骤而进入损伤识别位点，也就不会被切除，但 oxoG 会经历转换的最后步骤而进入损伤识别位点，从而被切除。

小结 碱基切除修复（BER）主要作用于损伤轻微的碱基。由 DNA 糖基化酶起始碱基切除修复过程，该酶使碱基对中的损伤碱基外突，然后切除损伤碱基，形成无嘌呤或无嘧啶位点，在 DNA 修复酶的作用下，切除磷酸戊糖，由正常的核苷酸取代。在细菌中，由 DNA 聚合酶 I 填补缺失的核苷酸；在真核生物中，由 DNA 聚合酶 β 完成填补任务。然而 DNA 聚合酶 β 易出现错误，且自身又无校正活性，因此由 APE1 执行校正功能。对 DNA 分子内 8-氧代鸟嘌呤位点的修复是 BER 的一个特殊事例，通过两个途径进行修复。由于 oxoG 会与 A 错配，所以，在 DNA 复制结束后由腺嘌呤 DNA 糖基化酶将 A 切除，然而，如果没有进行复制，oxoG 仍然与 C 配对，则由另一种 DNA 糖基化酶 oxoG 修复酶将 oxoG 切除。

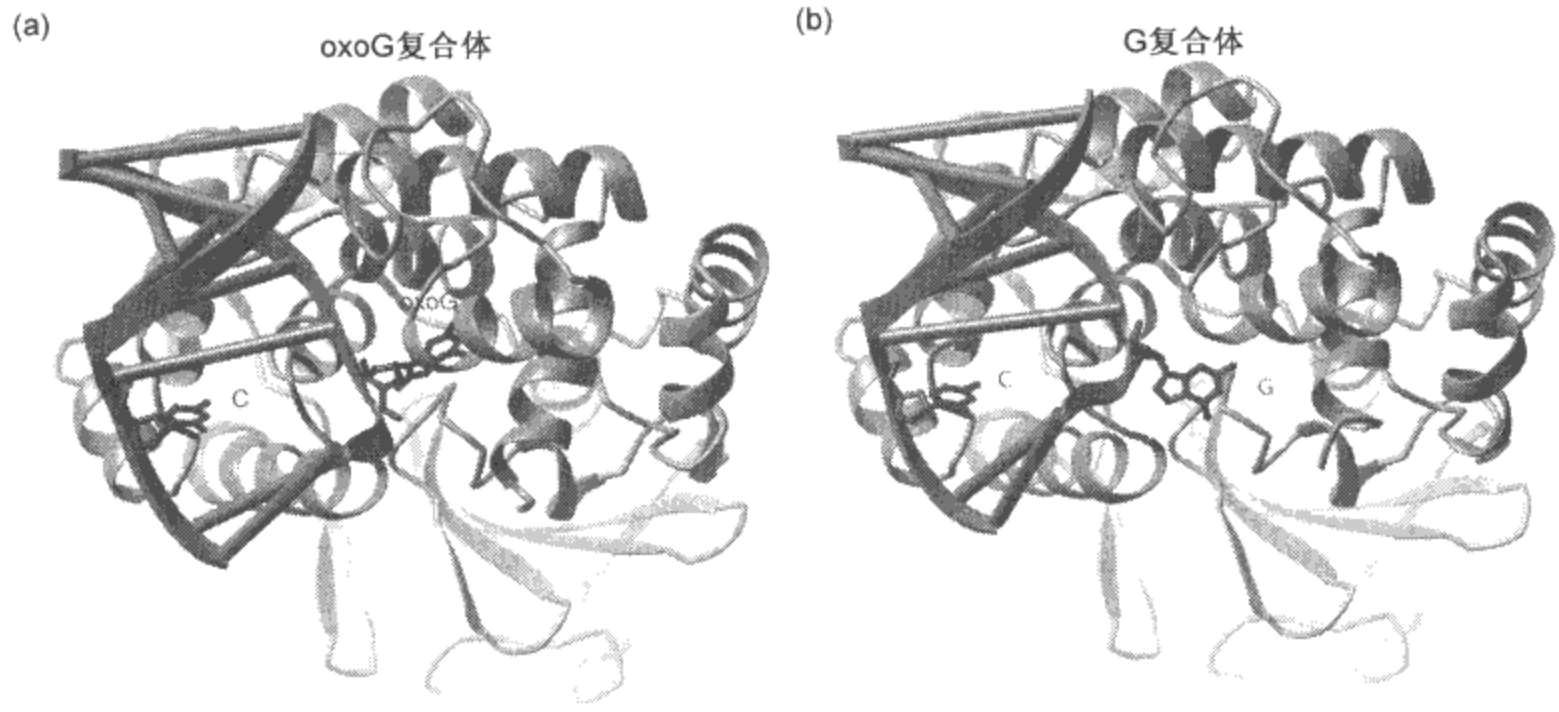


图 20.35　在包含有（a）oxoG-C 碱基对和（b）G-C 碱基对的复合体中，hOGG1 的晶体结构。蛋白质用蓝色表示，DNA 骨架用橙黄色表示，远离的 C 用紫红色表示，oxoG 和 G 用红色表示。（*Source*：(a-b) *Nature*, 434, Anirban Banerjee, Wei Yang, Martin Karplus and Gregory L. Verdine, "Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA," fig. 2, p. 614, copyright 2002, reprinted by permission from Macmillan Publishers Ltd.）

核苷酸切除修复　包括胸腺嘧啶二聚体在内，当碱基受到较大程度的损伤时，会被修复系统直接切除，且无须 DNA 糖基化酶的协助。在核苷酸切除修复（nucleotide excision repair，NER）途径中（图 20.36），切除酶系统识别发生严重损伤的 DNA 链，然后从损伤区域的两侧将 DNA 链断裂，切除包含损伤区域在内的一段寡核苷酸链。*E. coli* 细胞核苷酸切除修复途径中的关键酶是 uvrABC 内切核酸酶，该酶由 3 个多肽组成，分别由 *uvr*A、*uvr*B 和 *uvr* C 基因所编码。uvrABC 内切核酸酶依据受到损伤的是一个碱基（烷基化修饰）还是两个碱基（胸腺嘧啶二聚体），将受损伤的 DNA 链切割成长度为 12～13nt 的寡核苷酸片段。催化核苷酸切除修复反应的酶是切除核酸酶或切割酶（excinuclease）。我们即将了解到，在真核细胞中核酸酶切除的寡核苷酸片段的长度为 24～32nt，而不是 12 聚体或 13 聚体，但不管怎样，都是由 DNA 聚合酶填补空隙、由 DNA 连接酶将切口黏合。

有关人类 DNA 修复机制的信息主要来自对先天性 DNA 修复缺陷疾病的研究。这种 DNA 修复失调导致人类疾病的发生，包括 Cockayne 氏综合征和着色性干皮病（xeroderma pigmentosum，XP），当暴露在阳光下后，与正常人相比，大多数 XP 患者得皮肤癌的概率高数千倍。事实上，皮肤癌患者的皮肤只是在表面着生一些雀斑，然而，如果 XP 患者免受阳光曝晒，他们患皮肤癌的概率与正常人是相同的，即使将 XP 患者曝晒于阳光之下，未遭受阳光照射的皮肤几乎不会发生癌变，这些发现充分表明阳光是一种潜在的诱变剂。

为什么 XP 患者对阳光如此敏感？这是因为 XP 患者表皮细胞的 NER 系统发生缺陷，不能有效地修复 DNA 螺旋发生扭曲的损伤，包括嘧啶二聚体损伤。因此，得不到修复的损伤被保留下来并最终导致突变的发生，进而引发癌症。由于 NER 对化学诱导的 DNA 螺旋扭曲损伤也有修复作用，所以预测 XP 患者更易受化学诱变剂的影响而患癌症，事实也确实如此。然而，XP 患者患这些癌症的概率只是略高于正常人，这说明人类细胞内的大部分 DNA 损伤并不是螺旋扭曲，通过 BER 途径即可修复这些轻微的损伤。但是没有修复 UV 损伤的替代途径，因为人类没有光复活修复系统。

真核生物的核苷酸切除修复可通过两种途径完成对损伤的修复，包括可以修复基因组内的所有损伤的全基因组核苷酸切除修复（全基因组 NER 或 GG-NER）和只修复局限于转录链上的基因组活性区损伤的转录偶联核苷酸切除修复（transcription-coupled NER 或 TC-NER）。这两种形式的 NER 具有许多共同点，但它们识别损伤的方式不同，下面我们将学习发生在人类细胞内的这两种修复过程。

全基因组 NER　在 XP 患者的细胞内，修复系统的哪一个步骤发生了缺陷？关于这个问题的答案至少有 8 个。深入开展这个问题研究的途径是，将

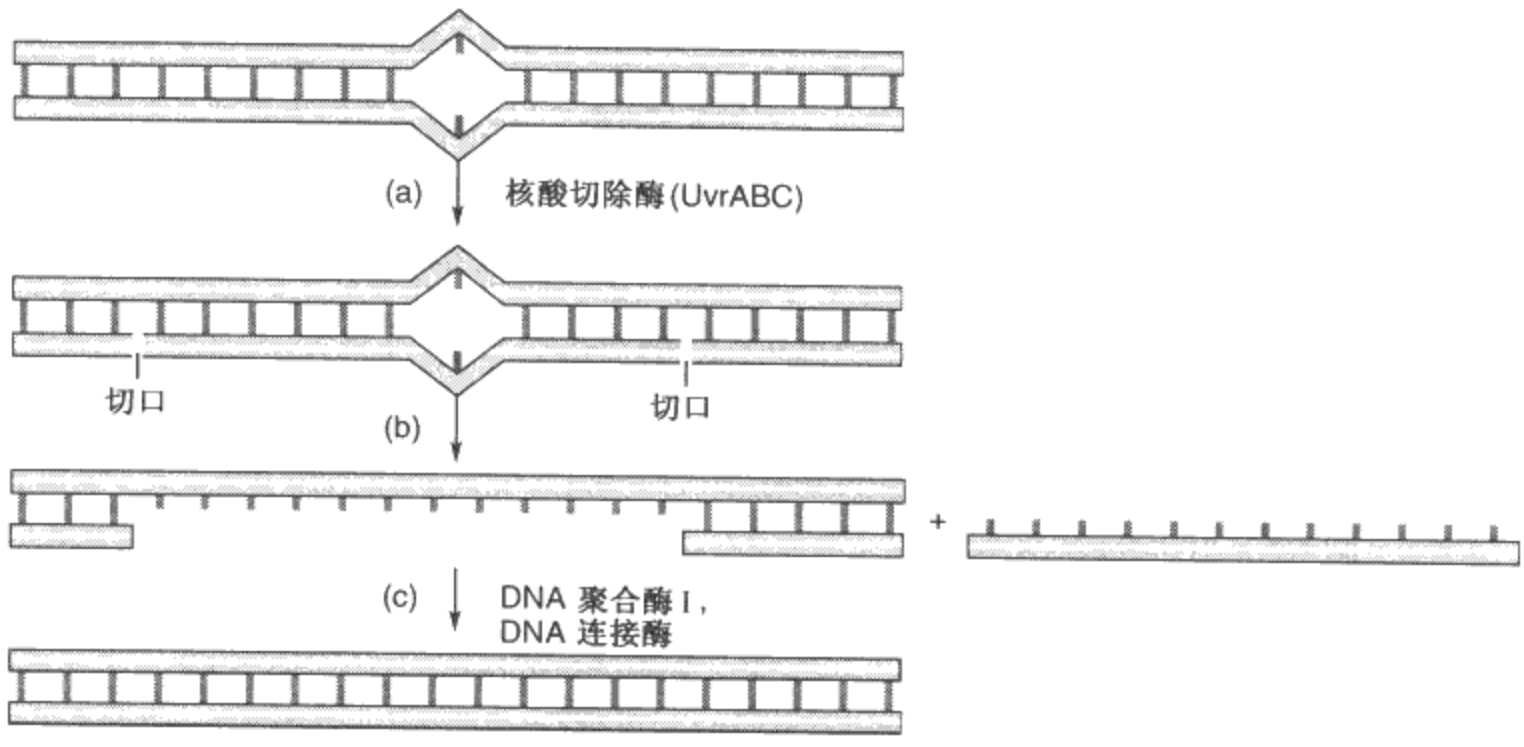


图 20.36　*E. coli*的核苷酸切除修复。（a）UvrABC 切除核酸酶在严重受损碱基（红色）的两侧对 DNA 链进行切割，切除（b）中一段长度为 12nt 的寡聚核苷酸片段。如果是嘧啶二聚体损伤，那么切除的寡聚核苷酸片段将为 13 聚体而不是 12 聚体。（c）DNA 聚合酶 I 以上方链为模板填补缺失的核苷酸，如同碱基切除修复，最后由 DNA 连接酶完成对切口的粘合。

来自不同患者的细胞进行融合，然后观察这些融合细胞是否还仍然具有缺陷症状。通常情况下，融合细胞是不会表现缺陷症状的，除非来自 2 个患者的基因彼此互补，这种可能性意味着不同的患者发生缺陷的基因是不相同的。到目前为止，通过细胞融合方式，鉴定了影响切除修复的 7 个互补群，此外，一些患者表现出变异的 XP（XP-V），其切除修复系统正常，且患者细胞对 UV 的敏感性略高于正常人细胞。在本章的后面内容中，我们将进一步探讨与 XP-V 有关的基因。综上所述，已有的研究表明，至少有 8 个基因与修复缺陷有关，其中的 7 个 *XPA-XPG* 基因与切除修复有关，常常是切除修复的第一步骤出现缺陷，即不能切断受损的 DNA 链。

人类全基因组 NER 的第一步（图 20.37）是识别因 DNA 损伤而引起的双螺旋扭曲。此步骤需要 XP 蛋白质（XPC）的参与，XPC 与 hHR23B 蛋白共同识别损伤的 DNA，与 DNA 结合后，使损伤周围的一小段 DNA 区域发生熔解。1977 年开展的体外研究支持了解旋 DNA 的功能，研究中使用的模板其损伤邻近或包围于小的“泡状”解旋 DNA 区域。这些模板的修复不需要 XPC，表明 DNA 的熔解也就意味着 XPC 蛋白功能的结束。1998 年，Jan Hoeijmakers 及同事用 DNA 酶足迹法证实，XPC 能够直接与 DNA 上的扭曲位点结合并引起 DNA 构象的改变（可能是链的分离）。

XPA 具有对损伤 DNA 的亲和性，因此也参与切除修复过程中早期阶段的损伤识别。既然 XPC 和 XPA 都能够与损伤 DNA 结合，那么有什么证据证明 XPC 是切除修复过程中的第一个作用因子？Hoeijmakers 及同事用大小不同的模板进行竞争性实验，实验结果支持了这种假设。研究者将 XPC 与一种损伤模板混合，同时将除 XPC 之外的其他因子与另一种损伤模板混合，然后将这两种混合体系混合在一起，结果发现，最初与 XPC 混合在一起的损伤模板优先获得修复，表明 XPC 最先与损伤 DNA 结合。那么，XPA 有何作用？它可以与 NER 体系中的其他许多因子结合，所以 XPA 的功能可能是验证变性 DNA 分子（XPC 或其他方式使 DNA 变性）上的损伤，然后帮助 NER 体系中的其他因子在损伤位点上募集。

首先，令人惊讶的是，两个 XP 基因 *XPB* 和 *XPD* 分别编码通用转录因子 TFIIH 的两个亚基，意味着此通用转录因子参与了 NER 过程。然而，现在了解到，这两个多肽具有 TFIIH 固有的 DNA 解旋酶活性（参见第 11 章），所以，TFIIH 的功能之一就是扩大 DNA 损伤位点周围的熔解区域，但是，体外实验发现，即使损伤 DNA 易发生较大区域的熔解，NER 系统仍然需要 TFIIH 的参与，表明该蛋白质除具有 DNA 解旋酶作用外，还发挥其他的作用。事实上，TFIIH 能够与许多其他的 NER 因子发生相互作用，说明 TFIIH 是 NER 复合体的组织者。

被 TFIIH 熔解的 DNA 可吸引核酸酶的结合，在受损链的损伤位点的两侧将 DNA 链切断，切除包

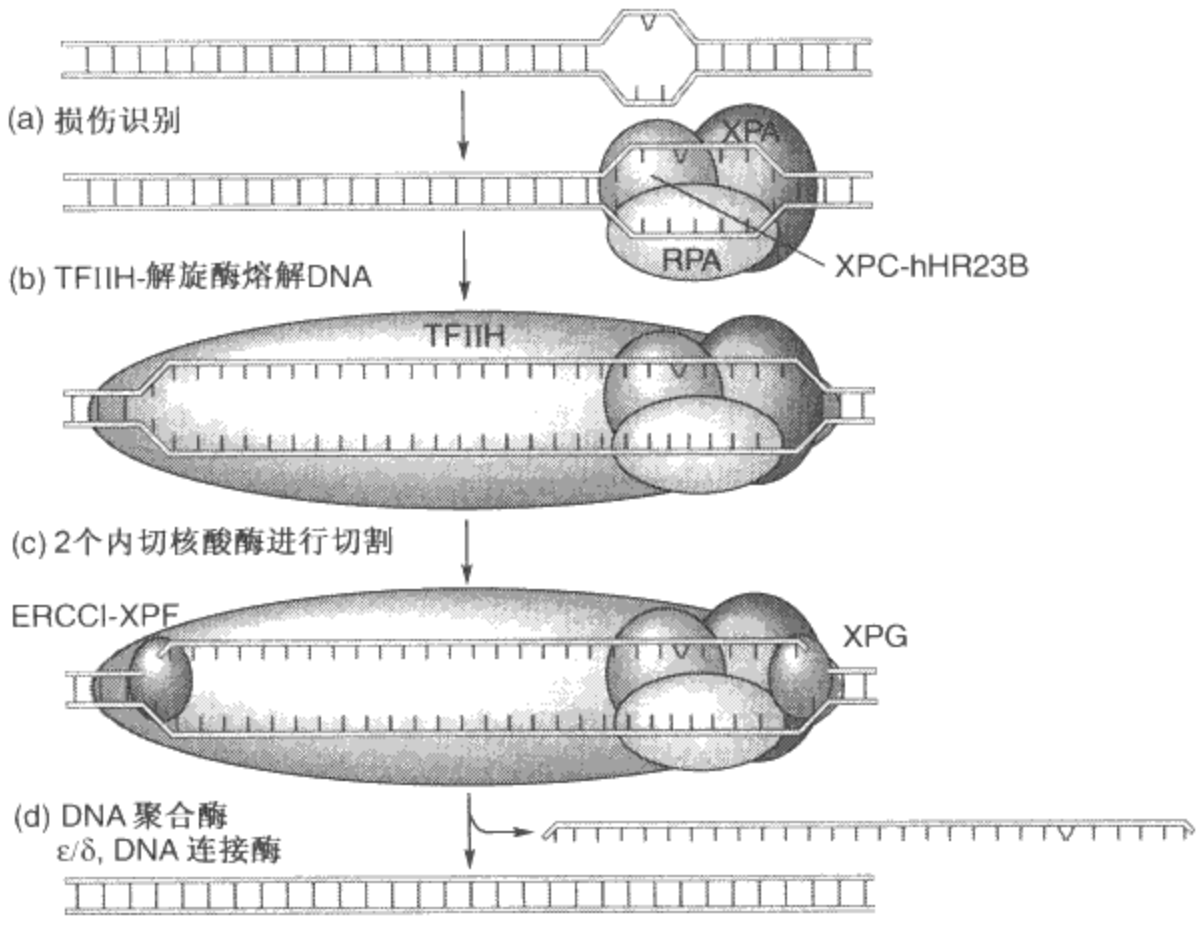


图 20.37　人类全基因组的核苷酸切除修复。(a) 在损伤识别阶段，XPC-hHR23B 复合体识别损伤部位（以嘧啶二聚体为例）并结合到该部位，导致 DNA 局部熔解，XPA 发挥辅助作用。RPA 与受损部位对面的完好 DNA 链结合。(b) TFIIH 的 DNA 解旋酶活性扩大了 DNA 的熔解区域。(c) RPA 协助 2 个内切核酸酶（ERCC1-XPF 复合体和 XPG）定位到损伤 DNA 的两侧，将 DNA 链切断。(d) 切除长度为 24～32nt、包括损伤位点在内的 DNA 片段，DNA 聚合酶填补缺口，最后由 DNA 连接酶将切口粘合。

含损伤位点在内的、长度为 24～32nt 的寡核苷酸片段。切除核酸酶 XPG 在损伤位点的 3′端一侧产生切口，ERCC1 与 XPF 组成的复合体在损伤位点的 5′端一侧产生切口，这两种核酸酶能够特异地切割双链 DNA 与单链 DNA 的连接点，且单链 DNA 是由 TFIIH 在受损位点处产生的。RPA 蛋白质能够帮助切除核酸酶准确定位，并进行正确切割。作为一种单链 DNA 结合蛋白，RPA 优先与受损部位对面完好的 DNA 链结合，RPA 面向完好 DNA 链 3′端的一侧与 ERCC1-XPF 结合，RPA 的另一侧与 XPG 结合，从而将这两个切除核酸酶定位在损伤位点两侧的正确位置上。

一旦损伤的 DNA 被切除，DNA 聚合酶 ε 或 δ 就会填补空隙，由 DNA 连接酶将切口粘合。有关 XPE 的功能目前还不清楚，尽管该蛋白质在体内具有重要功能。当 XPE 发生缺陷时，就会表现出 XP 症状，但在体外实验中，其功能不是必需的，所以很难确定 XPE 的具体功能。

转录偶联 NER　除 XPC 外，转录偶联 NER 中的其他因子与全基因组 NER 中的完全相同。在 GG-NER 体系中，XPC 的功能是起始损伤 DNA 的识别及局部 DNA 的熔解，那么在 TC-NER 体系中，哪一个因子具有这样的功能？当遇到因 DNA 损伤而引起的双螺旋扭曲时，RNA 聚合酶就会停止转录，在损伤位点，RNA 聚合酶促使 DNA 解旋形成熔解泡。XPA 识别变性 DNA 中的损伤位点，并募集其他因子。从此阶段开始，这些因子发挥在 GG-NER 中相同的功能，扩大 DNA 的熔解区域，在两个位点除切割 DNA，切除包含损伤位点在内的 DNA 片段。

RNA 聚合酶是高效实用的 DNA 损伤检测器，它能够在转录的同时持续扫描基因组，但遇到损伤位点时，RNA 聚合酶的转录过程就会终止，使 RNA 聚合酶对该损伤位点进行处理。当损伤位于非转录的 DNA 区域内（或即使在转录区的非转录链上）时，该损伤不会通过这种途径被发现。由于这些损伤不会影响基因的表达，所以稍后才能获得修复。事实上，非编码损伤如嘧啶二聚体、3mA 等可以像阻断 DNA 复制那样阻断转录，这对细胞而言是十分有益的，损伤导致正在转录的 RNA 聚合酶发生停止，从而募集损伤修复机构，实现损伤修复。

小结　核苷酸切除修复（NER）主要作用于会导致 DNA 双螺旋发生扭曲的严重损伤。*E. coli* 细胞进行核苷酸切除修复时，内切核酸酶在损伤位点两侧、相距 12～13nt 处将 DNA 链切断，从而切除长度为 12～13nt 的寡核苷酸片段。由 DNA

聚合酶 I 填补空隙、由 DNA 连接酶将切口粘合。真核生物的核苷酸切除修复通过两个途径完成对损伤的修复，在 GG-NER 途径中，XPC 与 hHR23 组成的复合体与基因组中的损伤位点结合，导致 DNA 的局部熔解从而起始修复过程。该蛋白质显然能够募集 XPA 和 RPA，随后 TFIIH 加入到复合体中，它的 2 个亚基（XPB 和 XPD）利用自身的 DNA 解旋酶活性延伸 DNA 的熔解区域。RPA 与 2 个切除核酸酶（XPF 和 XPG）结合，将这 2 个切除核酸酶定位在损伤位点两侧的正确位置上，然后释放出包含损伤位点在内的、长度为 24～32nt 的寡核苷酸片段。TC-NER 与 GG-NER 的机制十分相似，只是由 RNA 聚合酶发挥 XPC 的功能，检测损伤位点并起始 DNA 的熔解，在 TC-NER 与 GG-NER 途径中，一旦损伤的 DNA 被切除，DNA 聚合酶 ε 或 δ 就会填补空隙，由 DNA 连接酶将切口粘合。

真核生物的双链断裂修复

双链断裂可能是真核生物最严重的 DNA 损伤形式，如果得不到及时修复，可能会引起染色体的断裂，导致细胞的死亡或脊椎动物的癌症。真核生物可通过两种方式修复 DNA 的双链断裂（double-strand break，DSB）损伤：第一种修复方式是同源重组，以未断裂的姊妹染色单体作为重组片段的供体，其机制与原核生物的重组修复相似，只是 DNA 的两条链均参与了重组；第二修复方式是非同源末端连接（nonhomologous end-joining，NHEJ），在细胞分裂的 S 期和 G_2 期，同源重组是 DNA 损伤修复的主要机制，因为只有 1 个 DNA 拷贝发生断裂，另一个拷贝可用于准确地校正断裂区。分裂频繁的酵母细胞主要依靠同源重组修复双链断裂损伤。此外，处于 G_1 期的哺乳细胞主要采用非同源末端连接方式修复双链断裂损伤，本节主要讲述非同源末端连接的修复机制。

1994 年，J. Phillips 和 W. Morgan 将限制性内切核酸酶导入中国仓鼠卵巢细胞而研究非同源末端连接的机制。此限制性内切核酸酶能够对染色体的 DNA 双链进行切割，在腺嘌呤磷酸核糖转移酶（adenine phosphoribosyltransferase，APRT）基因内也存在一个酶切位点，该基因在中国仓鼠卵巢细胞内只有 1 个拷贝。研究者筛选 *APRT* 基因发生突变的存活细胞，并对突变的 *APRT* 基因进行测序，以探明在连接过程中所发生的事件。研究者发现，在酶切位点附近会插入或缺失很短的 DNA 片段，而且这些插入或缺失明显受存在于 DNA 末端、长度为 1～6bp 的微同源区所控制，图 20.38 描述了非同源末端连接模型，并阐释了相应的研究发现。

在进行非同源末端连接过程中，首先是 DNA 末端与异源二聚体蛋白 Ku 结合，其两个亚基分别为 Ku70（M_r=69kDa）和 Ku80（M_r=83kDa）。Ku 蛋白的一个重要功能就是保护 DNA 的末端，防止在末端连接完成之前被降解。Ku 蛋白具有 DNA 依赖的 ATP 酶活性，是 DNA 蛋白激酶（DNA protein kinase，DNA-PK）的调节亚基，而 DNA-PKcs 是 DNA-PK 的催化亚基。X 射线晶体学研究表明，像手指上套一圆环那样，Ku 结合在 DNA 末端，其 2 个亚基组成圆环结构，内部充满碱性氨基酸残基，从而使 Ku 与酸性 DNA 紧密结合。

一旦 Ku 完成与 DNA 末端的结合，便能够招募 DNA-PKcs 及其他蛋白质组装成完整的 DNA-PK 复合体。每个 DNA 末端都有该蛋白质复合体的结合位点，且结合位点不仅仅局限在 DNA 末端，而且延伸到邻近末端的双链 DNA 区域。通过与其他 DNA 片段的结合，DNA-PK 复合体能够促进联会或使微同源区排列在一起。

2 个 DNA-PK 复合体能够彼此进行磷酸化修饰，这种磷酸化作用可产生两种效应。首先，DNA-PKcs 磷酸化后能够促进其催化亚基的解离；其次，Ku 的磷酸化可以激活其自身的 DNA 解旋酶活性，使 DNA 末端发生解旋。通过末端解旋而促使微同源区进行碱基互补配对，未参与配对链的末端部分被甩出；最后，核酸酶将甩出的片段降解，DNA 聚合酶填补空隙，DNA 连接酶将两条未配对链连接在一起。

当 DNA 的侧翼序列被切除后，DNA 分子可能会丢失或增加几个核苷酸，但丢失或添加核苷酸的数目是不确定的。在第 23 章讨论抗体基因的同源重组时，我们还会再次学习非同源末端连接的内容。抗体基因在同源重组时，会有意在 DNA 分子内引入双链断裂，通过选择游离的 DNA 末端而实现 DNA 片段的重排，该过程需要 Ku 的参与。

小结 通过同源重组或非同源末端连接可修复双链 DNA 断裂损伤。非同源末端连接修复过程需要 Ku 和 DNA-PKcs 的参与，均与 DNA 末端结合，组装成有活性的 DNA-PK 复合体，使 DNA 末端的微同源区彼此排列在一起，然后 2 个 DNA-PK 复合体彼此进行磷酸化修饰，这种磷酸化作用可以激活 DNA-PKcs 催化亚基的解离并激活 Ku 的 DNA 解旋酶活性，使 DNA 末端发生解旋。末端解旋后，微同源区进行碱基互补配对。最后额外的侧翼 DNA 片段被切除，空隙填补后，DNA 的末端被永久连接在一起。

染色质重建在双链断裂修复中的作用 在第 13 章我们曾学习到，核小体能够阻断基因的控制区与

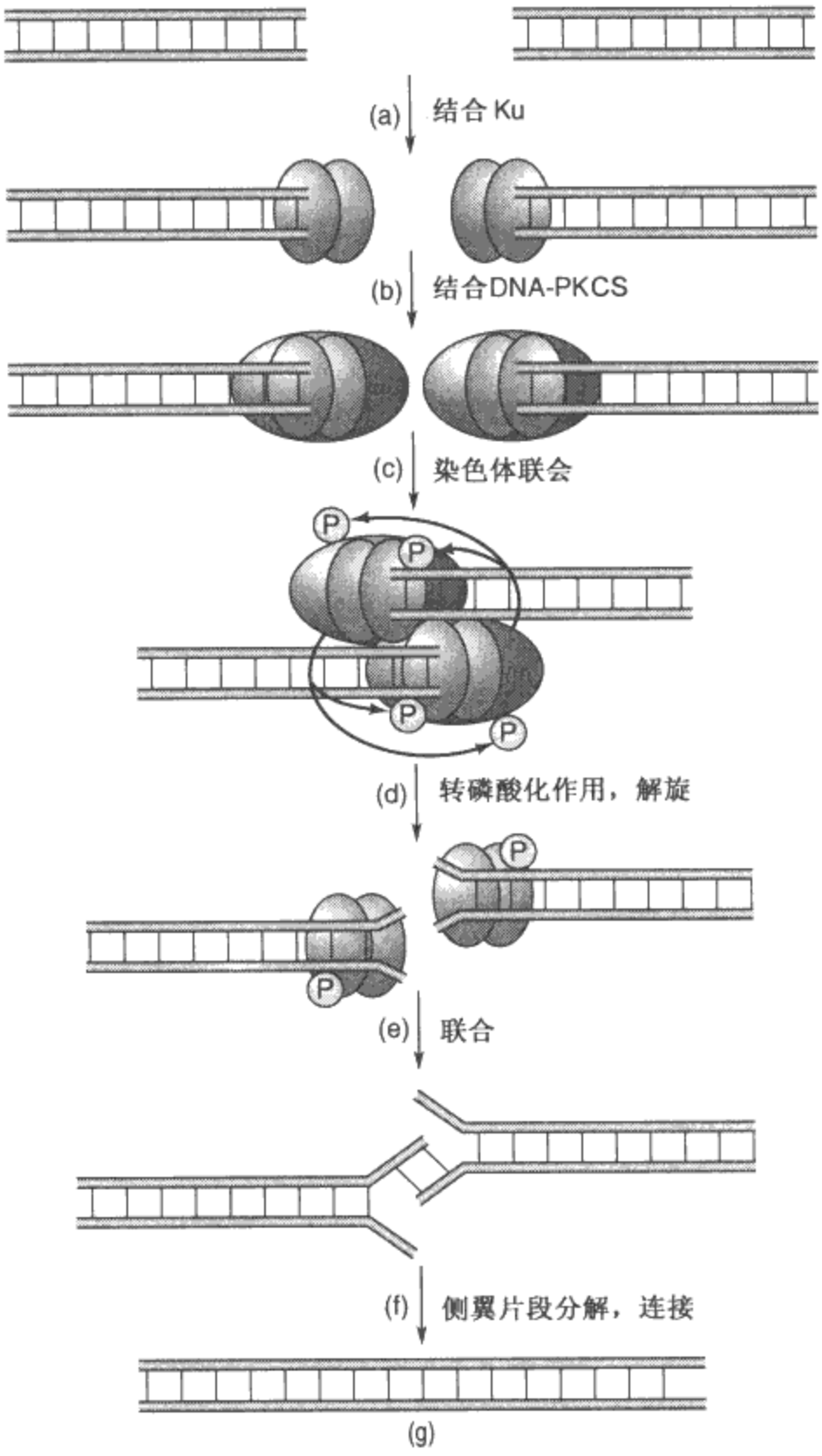


图 20.38　非同源末端连接模型。（a）游离的 DNA 末端与 Ku（蓝色）结合，起保护功能，防止 DNA 末端降解；（b）Ku 募集 DNA-PKcs（红色）的结合，组成完整的 DNA-PK 复合体；（c）DNA-PK 复合体促进联会，或将邻近 DNA 末端的微同源区域排列在一起；（d）两个 DNA-PK 复合体彼此催化复合体中的调节亚基（Ku）和催化亚基发生磷酸化修饰，然后，磷酸化的催化亚基与复合体解离；（e）Ku 的磷酸化修饰激活其自身的 DNA 解旋酶活性，使 DNA 的两端发生解旋；（f）两端的微同源区在联合阶段彼此配对；（g）侧翼的 DNA 片段被清除，填补缺口。最后，DNA 连接酶将 DNA 链的末端连接起来。（*Source*：Adapted from Chu，G.，Double strand break repair. *Journal of Biological Chemistry* 272 [1997] p. 24099，f. 4.）

转录因子之间的联系，只有通过染色质重建才能激活真核生物的基因。出于同样的原因，可以推测核小体也能够阻断损伤 DNA 与修复因子之间的联系。因此，损伤 DNA 的修复也必须要进行染色质重建。2004 年，Susan Gasser 及同事、Xuetong Shen 及同事这两个研究小组通过研究表明，在酵母的双链染色体断裂修复过程中，主要是通过同源重组进行修复的，而在修复过程中确实依赖于染色质重建复合体 INO80 的参与。

INO80 是染色质重建蛋白质 SWI/SNF 家族（参见第 13 章）中的成员，由 12 个亚基组成，包括 *ino*80 基因的产物 Ino80，此肽链具有染色质重建蛋白质特有的 ATP 酶/转座酶结构域。*ino*80 基因突变能够同时阻断基因的转录和 DSB 修复，可能的原因是染色质重建发生缺陷。

上述 2 个研究小组在酵母染色体的 *MAT* 基因座的特定位点处断裂 DNA 双链，然后用染色质免疫共沉淀（Chromatin immunoprecipitation，ChIP，参见第 13 章）方法计量结合在断裂处的蛋白质，在 30～60min 内 INO80 与断裂处结合，表明 INO80 参与了 DSB 修复过程。下一个值得关注的问题是，哪些蛋白质参与了对 INO80 的募集？线索之一是酵母蛋白激酶 Mec1 和 Tel1 能够对邻近 DSB 处核小体组蛋白 H2A 的第 129 位丝氨酸残基进行磷酸化修饰，用丙氨酸残基取代第 129 位丝氨酸残基后，酵母细胞对能够造成 DNA 损伤的射线及化学试剂十分敏感。因为，与丝氨酸不同，丙氨酸不能被磷酸化修饰。这一发现表明，组蛋白 H2A 第 129 位丝氨酸残

基的磷酸化修饰能够促进 DSB 修复。

而且，编码 Mec1 和 Tel1 的基因突变或发生 129 位丝氨酸残基被丙氨酸残基取代的突变都能够阻止 INO80 在 DSB 位点的募集。这些发现说明磷酸化的 H2A 与 INO80 之间存在着直接的相互作用。Shen 及同事研究发现，INO80 确实能够与磷酸化的 H2A 和其他组蛋白共纯化，而与非磷酸化的 H2A 之间则无共纯化现象。

在 DSB 修复过程中，INO80 发挥着怎样的作用呢？Gasser 及同事研究发现编码 INO80 亚基的基因发生突变的酵母菌株或组蛋白 H2A 第 129 位丝氨酸发生改变的突变酵母菌株在断裂染色体的末端都不会形成外突的单链 3′端。所以，INO80 的功能之一就是在断裂处形成外突的单链 3′端，通过使核小体滑离断裂末端而利于外突的单链 3′端的形成。

INO80 的两个亚基 Rvb1 和 Rvb2 具有 ATP 酶活性，如同*E. coli*中参与重组和 DSB 修复的 RuvB 蛋白，RuvB 是由两个环形六聚体组成的同源多聚体蛋白（参见第 22 章），利用自身的 DNA 解旋酶活性而推动“分支位移”，一个连接两条重组 DNA 双链的分支的滑动。同样，Rvb1/Rvb2 也由双六聚体构成，具有 DNA 解旋酶活性，但还不清楚否具有推动分支滑动的功能。由于 DNA 解旋酶在对 DNA 解旋时，能够沿着 DNA 分子滑动，因此，我们也可以想象，INO80 利用自身 DNA 解旋酶活性，在沿 DNA 分子移动时能够拖动核小体，使其远离 DSB。

染色质重建蛋白 SWR1 也能被募集到 DSB 处，同 INO80 一样，SWR1 也包含有 Rvb1 和 Rvb2，而且还有一个令人感兴趣的活性：能够用 H2A 的变异体 Htz1 取代 H2A。所以，SWR1 的作用可能是用不能被磷酸化修饰的 Htz1 取代磷酸化的 H2A。如此可见，一旦 DSB 修复开始并处于修复过程中时，SWR1 就能够将核小体的磷酸化组蛋白回复到断裂时的状态。Jerry Workman 及同事通过研究支持了这一假说，研究发现在体外，果蝇的 SWR1 同源蛋白 Domino/p400 能够用非磷酸化的 H2A 取代磷酸化的 H2A。

小结 蛋白激酶 Mec1 和 Tel1 被募集到 DSB 位点后，使核小体组蛋白 H2A 的第 129 位丝氨酸残基发生磷酸化修饰，而这种磷酸化作用能够使染色质重建蛋白 INO80 被招募到 DSB 处，该蛋白质利用自身 DNA 解旋酶活性推动核小体远离 DSB 的末端，导致在断裂处形成外突的单链 DNA 3′端，该末端是非同源末端连接和同源重组所必需的。另一种染色质重建蛋白 SWR1 与 INO80 共享多种组分，并也被募集到 DSB 处，可用无磷酸化修饰的 H2A 变异体 Htz1 取代 H2A。使邻近 DSB 位点处已发生磷酸化的 H2A 恢复正常状态。

错配修复

到目前为止，我们已经讨论了有关因化学诱变剂造成 DNA 损伤的修复机制，但对于只是由于掺入错误碱基且被校正系统遗漏的错配又将如何获得修复？首先，修复这样的错误十分棘手，因为很难确定哪一条是带有错误核苷酸的新合成链，哪一条是不需要修复的正确的亲本链。在*E. coli* 细胞中这个问题很容易解决，因为亲本链具有区别于子链的识别标签，这些标签就是甲基化的腺嘌呤。甲基化酶可以识别 5′-GATC-3′ 序列，然后将甲基转移到腺嘌呤上。该 4 碱基序列广泛分布于整个基因组中，通常每隔 250bp 就会出现一次，邻近于可能出现错配的位点。

GATC 为回文序列结构，其互补链 5′→3′ 方向的序列也是 GATC。当母链甲基化的 GATC 序列被复制后，新合成子链上的 GATC 序列也将发生甲基化修饰，只是甲基化作用会有一定时间的延迟，使子代 DNA 呈半甲基化状态。错配修复系统（图 20.39）便会利用这种时间延迟，以亲本链上的甲基为标记，将亲本链和子链区分开来，只对未甲基化子链上的错配进行修复。该修复过程必须在错配发生后不久进行，否则子链也完成甲基化修饰，就无法区分亲本链和子链了。对真核生物的错配修复系统了解的还不是很清楚，编码错配识别蛋白质和切除酶（MutS 和 MutL）的基因都是十分保守的，所以，依赖这些酶的错配修复机制在真核与原核生物中是相似的。然而，在真核生物内并没有发现编码链识别蛋白（MutH）的基因，可能真核生物的错配修复系统并不采用甲基识别策略。现在还不清楚在错配发生处，真核生物细胞如何区分亲本链和子链。

小结 *E. coli* 的错配修复系统通过识别 GATC 序列中的甲基化腺嘌呤而将亲本链与子链区分开来，对亲本链的互补链（子链）上的错配进行修复。真核生物利用错配修复系统的部分功能，但其区分亲本链和子链的策略还不清楚。

人类细胞错配修复系统的失效

人类细胞错配修复系统失效后会产生严重后果，最为普遍的后果之一就是引发遗传性癌症。例如，遗传性非息肉结肠癌（hereditary nonpolyposis colon cancer，HNPCC），又称 Lynch 综合征，大约每 200 个美国人中就有 1 人患有这种癌症，占所有结肠癌的 15%。HNPCC 患者的特征之一就是微卫星序列不稳定性，也就是说，长度为 1～4bp 的串联重复序列，即

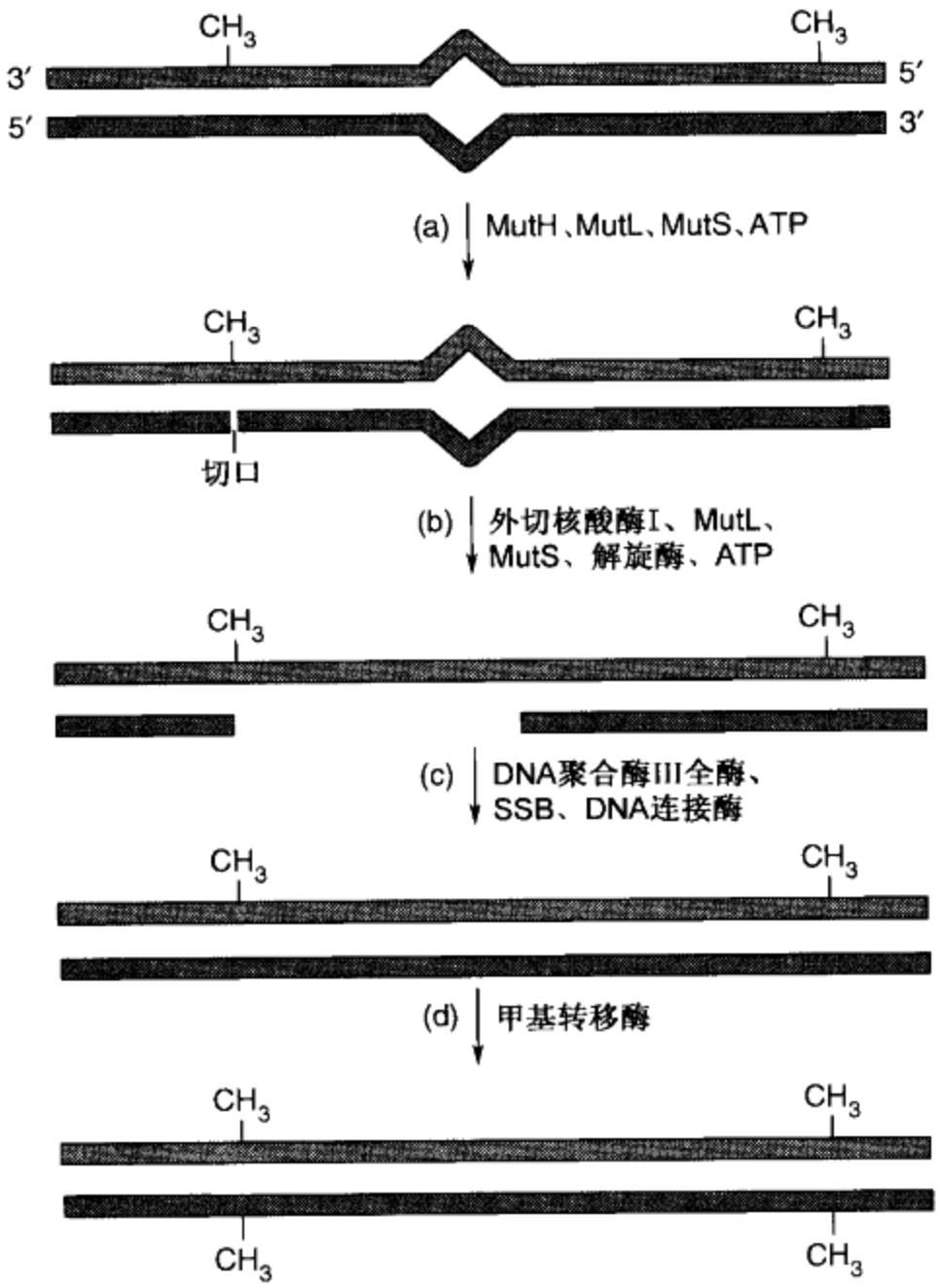


图 20.39　*E. coli*的错配修复。（a）*mutH*、*mutL* 和 *mutS* 基因产物与 ATP 一起识别错配碱基（中心），依据 GATC 序列缺少甲基基团而识别出新生子代 DNA 链，在甲基化 GATC 序列对面新链的错误核苷酸的上游引入一个切口；（b）外切核酸酶 I 与 MutL、MutS、DNA 解旋酶以及 ATP 一起将切口下游包括错配核苷酸在内的一段 DNA 切除；（c）DNA 聚合酶 III 的全酶在单链结合蛋白（SSB）的协助下，填补由外切核酸酶产生的缺口，DNA 连接酶将留存的切口粘合；（d）甲基转移酶将子链的 GATC 序列进行甲基化修饰。甲基化作用一旦完成，就不能再发生错配修复，因为此时已经无法区分子链和亲本链。

DNA 微卫星序列在患者的一生中会改变其大小（重复的数目），这是很不正常的现象。就某一给定的微卫星序列而言，其重复数目可能存在着个体差异，但在每个个体的所有组织中应该是恒定的，且在个体的一生中保持不变。微卫星序列不稳定性与错配修复系统之间的关系是在 DNA 复制期间因“打滑”而引起短的重复序列插入过多或过少，导致“泡状结构”的产生，错配修复系统能够识别并修复这些“泡状结构”。当错配修复系统出现故障时，DNA 复制期间的“打滑”就不会获得校正，这样，在细胞分裂进行 DNA 复制使就会导致许多基因发生突变，这种遗传不稳定性可能会引发癌症，特别是控制细胞分裂的基因（癌基因和肿瘤抑制基因）发生突变。

小结　人类细胞错配修复系统的失效会导致微卫星序列的不稳定性，最终会引发癌症。

未修复 DNA 损伤的处理

直接逆转修复和切除修复机制都是真正的修复过程，能够完全消除 DNA 缺陷。然而，细胞还存在处理损伤的其他方式，即不需要消除损伤而是简单地避开损伤，即使并不是真正地修复 DNA 损伤，但这种方式有时也被称为修复机制，损伤旁路机制或许是更为科学的术语。当细胞没有对损伤进行真正修复、且在损伤修复前已完成了 DNA 复制或已完成了 DNA 复制并进行了细胞分裂时，损伤旁路机制才会发挥作用，在 DNA 复制和细胞分裂过程中，细胞失去了处理 DNA 损伤的最佳机会，将会面临更大的危险。

重组修复　重组修复（recombination repair）是最为重要的旁路修复机制，有时也被称为复制后修复。当复制跳过嘧啶二聚体损伤位点时，子链上的缺口必须被修复，这种修复就是复制后修复。此时切除修复是无效的，因为在子链上对应于嘧啶二聚体的位置处无完好的 DNA，只是存在一个缺口，所以，重组修复是为数不多的替代途径之一。图 20.40 描述了重组修复的作用机制。首先，DNA 被复制，嘧啶二聚体会阻碍损伤位点的模板作用，使复制机构在嘧啶二聚体位点处停止，然而暂停之后，复制机构会跳过该位点继续进行复制（重新起始 DNA 复制可能需要合成新的引物），从而在子链上留下一个缺口。接着，具有缺口的 DNA 链与另一个子代双链 DNA 分子中的同源单链进行重组，重

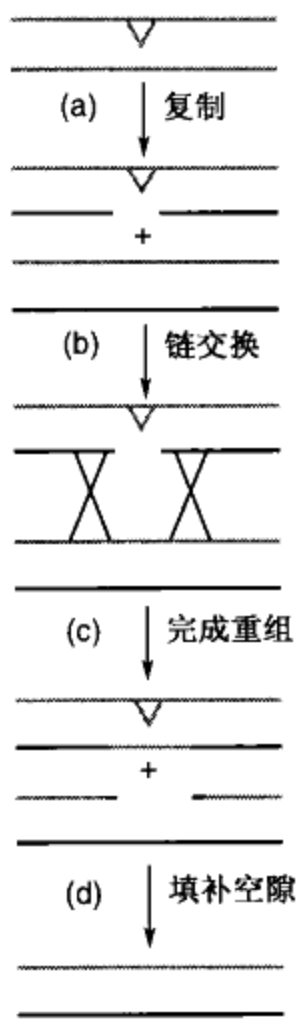


图 20.40　重组修复。用“V”形表示 DNA 分子上的嘧啶二聚体。(a) 在 DNA 复制时，复制机构会跳二聚体所在的区域，在该处留下一个缺口，而互补链正常复制，两条新合成的子链用红色表示。(b) 同源链之间发生链交换。(c) 重组完成后，嘧啶二聚体的缺口完成填补，但在另一子代双链 DNA 分子中留下了一个缺口。带有嘧啶二聚体的双链 DNA 并没有被修复，但成功地完成了复制，可能会在下一轮 DNA 复制时得到准确的修复。(d) 以正常的互补链作为模板，最后的缺口很容易获得填补。

组需要 *recA* 基因产物的参与，该蛋白质能够促进同源 DNA 链间的交换（参见第 22 章）。在讨论诱导 λ 原噬菌体的 SOS 反应时，我们对 *recA* 基因有所了解（参见第 8 章），在第 22 章学习重组时将获得更为全面的了解。重组后的实际效果是填补了子链上对应于嘧啶二聚体的缺口，而在另一个子代双链 DNA 分子上产生新的缺口，然而，由于该双链 DNA 分子上无二聚体，产生的缺口很容易被 DNA 聚合酶和连接酶所填补。值得注意的是 DNA 损伤依然存在，但是细胞至少可以进行 DNA 的复制，迟早会通过真正的 DNA 修复机制修复存在的 DNA 损伤。

易错旁路　所谓的易错旁路（error-prone bypass）是另一条处理 DNA 损伤的途径，在该途径中，DNA 损伤并没有获得真正的修复。在*E. coli*细胞中，作为 SOS 应答的一部分，该途径可被 DNA 损伤所诱导，包括 UV 损伤，而且依赖于 *recA* 基因产物的参与。具体的反应步骤如下（图 20.41）：UV 或其他诱变处理以某种方式激活 RecA 辅蛋白酶活性，该辅蛋白酶具有多个靶点，我们曾学习过的 λ 阻遏蛋白就是其中之一，但 *Lex*A 基因的产物 LexA 是其主要的靶点，LexA 是许多基因包括修复系统基因的抑制子，当 LexA 被 RecA 辅蛋白酶激活后，LexA 会自我分解，受其抑制的基因将被诱导表达。

被诱导的基因 *umuC* 和 *umuD* 组成一个操纵元 *umuDC*。*umuD* 基因的产物 UmuD 可被蛋白酶切割产生 UmuD′，与 *umu C* 基因的产物 UmuC 组成具有 DNA 聚合酶活性的 $UmuD'_2C$ 复合体，所以该复合体也被认为是 DNA pol V。在体外，pol V 自身能够引发 DNA 损伤处理的易错旁路，但 pol III 全酶和 RecA 蛋白对这个过程具有很大促进作用。这类旁路途径即使不可能正确阅读 DNA 损伤位点，也会完成对损伤位点的复制，从而避免了缺口的产生，但常常会在新合成的 DNA 子链中掺入错误的碱基（因此得名为“易错”）。当 DNA 再次复制时，这些错误得以保持。存在于真核生物内的易错旁路及无差错旁路机制通称为跨损伤合成（translesion synthesis，TLS）。

DNA pol V 能够有效地跨过 3 类最为普遍的 DNA 损伤：嘧啶二聚体；UV 引起的相关损伤如 [6-4] 光化产物（[6-4] photoproduct），即一个嘧啶碱基的 6-位碳原子与相邻嘧啶碱基的 4-位碳原子发生共价连接；无碱基的 AP 位点。然而，DNA pol V 在进行跨损伤合成时，其准确度存在一定的变化。2000 年，Myron Goodman 及同事测定了 DNA pol V 在跨损伤如胸腺嘧啶二聚体或 [6-4] 光化产物和 AP 位点合成时 A 和 G 的掺入量。在对应于胸腺嘧啶二聚体的位置处，DNA pol V 倾向于掺入 2 个 A，但如果二聚体中包含有胞嘧啶，情况就会出现变化。对应于由两个胸腺嘧啶形成的 [6-4] 光化产物位置处，DNA pol V 倾向于在第一个位点掺入 G、在第二个位点掺入 A，很显然，这不是十分保真的复制。在对应于 AP 位点的位置处，DNA pol V 掺入 A 的概率为 2/3、掺入 G 的概率为 1/3。碱基掺入比例以及没有检测到有嘧啶碱基掺入的事实与在体内观察到的情况相一致，表明 DNA pol V 确实是体内执行跨损伤合成功能的聚合酶。

如果 *umu* 基因确实负责易错旁路，那么可以预测，其中的一个 *umu* 基因发生突变将会使*E. coli* 细胞降低对突变的敏感性。这些突变细胞仍然会以相同的概率发生 DNA 损伤，但这些损伤不易被转变为

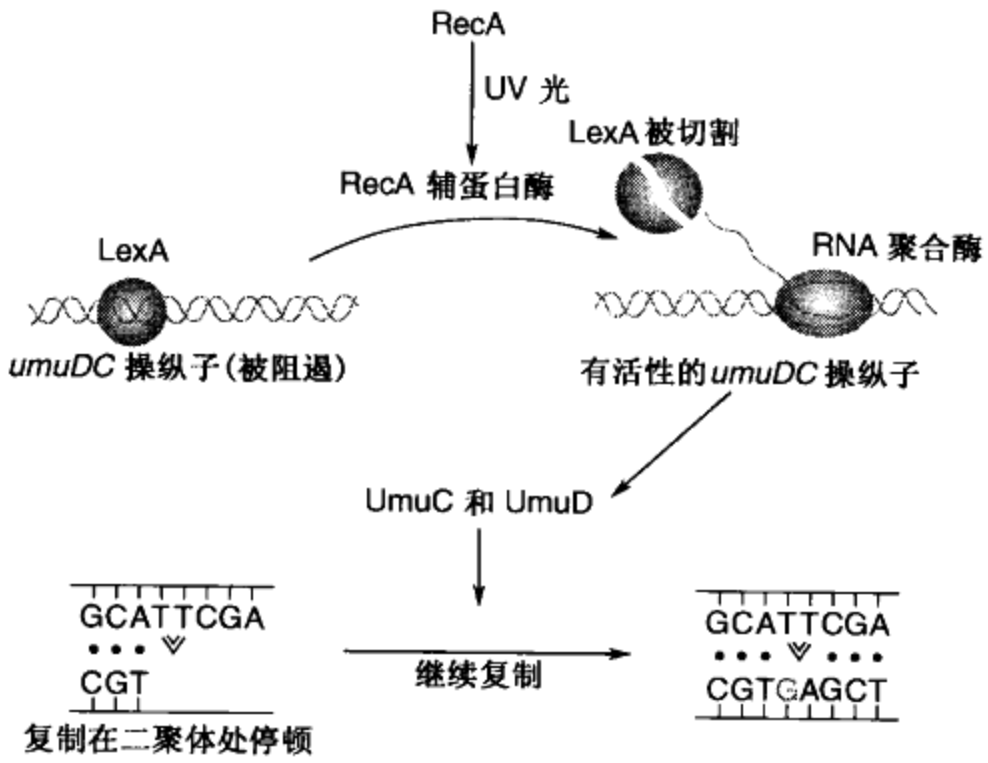


图 20.41　易错（SOS）旁路。　紫外线激活 RecA 辅蛋白酶，RecA 辅蛋白酶激活 LexA 蛋白（紫色），LexA 蛋白进行自我切割，使其从 *umuDC* 操纵元上释放出来。这一过程导致 umuD 和 umuC 蛋白的合成，这些蛋白质即使在经常出错（蓝色）的情况下，也能完成对嘧啶二聚体损伤位点的复制。

突变。1981 年，Graham Walker 及同事通过构建 *umu C* 的无效等位基因（null allele）（无活性基因）证实了这一推测，而且携带无效等位基因的细菌基本上不会发生突变，事实上，“umu”代表“无突变”。

研究者构建了一个携带 *umu C* 突变的*E. coli* 菌株，且带有一个 his^- 赭石突变，该突变通常可被 UV 照射所回复。研究者用 UV 照射该菌株，并计数 his^+ 回复突变体的数量，回复突变体越多，发生的突变也就越多，因为逆转就是一种回复突变。图 20.42 显示了研究结果，在野生型细胞中存在一定数量的回复突变体（在最高 UV 剂量时约有 200 个），与之形成鲜明对比的是，*umu* C^- 细胞几乎无回复突变的发生。此外，添加携带 *umu C* 基因的质粒后，能够抑制 *umu* C^- 细胞的无突变表型，导致回复突变体的数量显著增加（即使在相对较低的 UV 剂量时，回复突变体的数量也可达到 500 个）。

实验中使用的无效等位基因是通过如下途径获得的：将无 *lac* 启动子的 *lac* 结构基因插入到 *umu C* 内，然后筛选 lac^+ 细胞。因为细胞最初为 lac^- 表型，之所以出现 lac^+ 表型，是因为 *lac* 基因插入到 *umuDC* 启动子的下游。由于 *lac* 基因受到 *umuDC* 启动子的控制，因此，Walker 及同事通过检测 β-半乳糖苷酶的活性，即可测定 UV 照射对 *umuDC* 启动子的诱导性。图 20.43 显示，*umuDC* 启动子确实能够被剂量为 10J/m² 的 UV 照射所诱导（蓝色曲线），但是在 *lex*A 突变体或 *recA*$^-$ 细胞内，该启动子不受 UV 照射所诱导（红色曲线）。实验中所使用

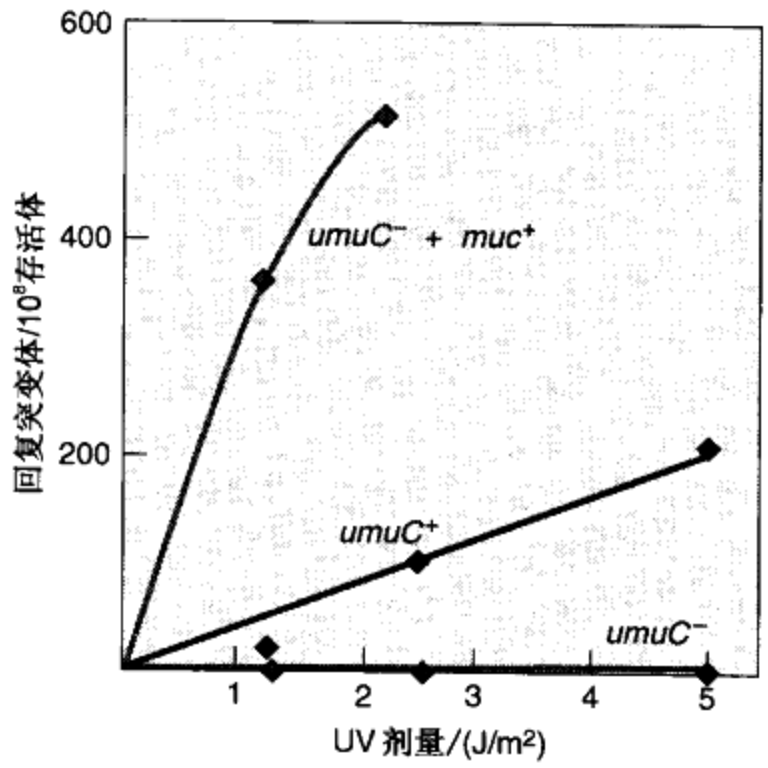


图 20.42　*E. coli umuC* 菌株的不易突变性。　Walker 及同事用紫外光辐射*E. coli*的 3 个 his^- 菌株，并检测其形成 his^+ 回复突变的能力。这 3 个菌株分别是：*umuC* 基因为野生型的菌株（蓝色）、*umuC*$^-$ 菌株（红色）及含有携带 *umuC* 基因的质粒的 *umuC*$^-$ 菌株（绿色）。（*Source*：Adapted from Bagg，A.，C. J. Kenyon and G. C. Walker，Inducibility of a gene product required for UV and chemical mutagenesis in *Escherichia. coli*. *Proceedings of the National Academy of Sciences USA* 78：5750，1981.）

的 *lexA* 突变体细胞编码的 LexA 蛋白不能裂解，所以不能将其从 *umuDC* 启动子上移除。

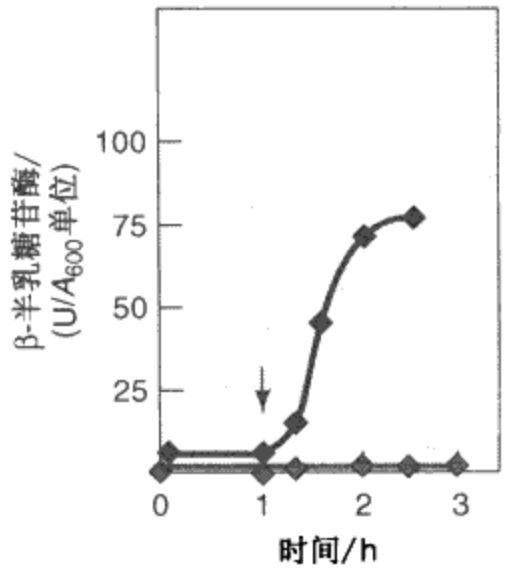


图 20.43　*umuDC* 启动子可被 UV 所诱导。Walker 及同事用剂量为 10J/m² 的 UV 照射携带受 *umuDC* 启动子调控的 *lac* 基因的细胞。如箭头所示，照射 1h 后测量每 OD_{600} 单位（用于度量细胞密度的混浊度指标）细胞的 β-半乳糖苷酶活性（黄色曲线）。研究者也对 *lexA* 突变株（绿色）和 *recA*⁻ 细胞（红色）进行了同样的实验。*lexA* 突变株是不可诱导的，突变的 *lexA* 基因编码的 LexA 蛋白不能被裂解，所以不能将其从 *umuDC* 启动子上移除。（*Source*：Adapted from Bagg，A.，C. J. Kenyon and G. C. Walker，Inducibility of a gene product required for UV and chemical mutagenesis in *Escherichia coli*. *Proceedings of the National Academy of Sciences USA* 78：5751，1981.）

野生型*E. coli* 细胞由于存在着有效的修复机制，所以能够忍受基因组中出现的多达 50 个嘧啶二聚体，而且不会产生不良影响。缺失 *uvr* 基因的菌株无法进行切除修复，因此对 UV 损伤极其敏感。然而，这些菌株对 DNA 损伤仍有一定程度的抗性。另一方面，*uvr* 和 *recA* 双突变菌株既不能进行切除修复，也不能进行重组修复，对 UV 损伤十分敏感，可能的原因是这类菌株只能通过易错修复途径进行 DNA 损伤修复。基因组内的 1～2 个嘧啶二聚体即可成为细胞的致死剂量。

很显然，如果细菌细胞在进化过程中没有获得易错旁路，它们将很少发生变异。如果事实如此的话，为什么还要保留这种导致突变的机制？这可能是由于易错旁路系统对细菌而言，其利大于弊，因为即使具有发生突变的危险，易错旁路系统也可以允许有机体复制受损的基因组。有一点是显而易见的，如果因复制失败而要付出细胞死亡的代价，细胞宁愿复制未经修复的损伤 DNA 并进行细胞分裂。这一系列过程中如果产生一个带有 DNA 缺口的子代细胞，此时，切除修复和重组修复都不能发挥作用，从而可能导致细胞的死亡。所以，最终的解决办法就是通过易错旁路系统来避免细胞的死亡。

一定程度的变异对一个物种而言是有利的，这种变异使生物体种群的基因组呈现趋异性，从而使生物个体间对疾病或其他侵害表现出不同的敏感性。当面临新环境的考验时，种群中的某些成员因具有抵抗力而存活下来，从而使种族得以延续。

小结　细胞可以通过非修复方式来处理 DNA 损伤，方式之一为重组修复，具有缺口的 DNA 链与另一个子代双链 DNA 分子中的同源单链进行重组，使缺口问题得以解决，但最初的损伤未获得修复。另一种处理损伤的机制是引发 SOS 应答，至少在*E. coli* 细胞中如此，该机制即使不可能正确阅读 DNA 损伤区域，也会完成 DNA 复制，会导致新合成 DNA 链上出现错误，因此这种处理 DNA 损伤的机制被称为易错旁路。

人类细胞的易错修复与无差错修复　在整个生命王国中，所有的 DNA 修复过程都是十分保守的，或许是因为在生命的开始之处，DNA 损伤就成了生命的一部分，所以 DNA 损伤修复机制的进化要早于生物的分化。易错旁路机制也不例外，人类细胞对 DNA 损伤如嘧啶二聚体的处理系统与原核生物的损伤处理系统类似，这些旁路系统有赖于特殊的 DNA 聚合酶，包括 DNA 聚合酶 ζ、DNA 聚合酶 η、DNA 聚合酶 θ、DNA 聚合酶 ι 和 DNA 聚合酶 κ。DNA 聚合酶 δ 和 e 负责前导链的合成，当遇到无结构的 DNA 损伤如嘧啶二聚体时，就会发生停顿，此时 DNA 聚合酶 δ 和 e 就会被上述特殊的 DNA 聚合酶所取代。

这些 DNA 聚合酶中某些成员在跨过 DNA 损伤位点时会随机地插入碱基，很显然，这是一种易错策略。但有些聚合酶的出错概率很低，是一种相对无差错的损伤处理策略。例如，DNA 聚合酶 η 会自动地在子链对应于嘧啶二聚体的位置处插入 2 个 dAMP，即使不能与嘧啶二聚体中的碱基形成互补的碱基对，只要二聚体中的碱基是胸腺嘧啶，该系统就能够做出正确的选择，而胸腺嘧啶二聚体又是嘧啶二聚体最为常见的形式。

1999 年，Fumio Hanaoka 及同事发现携带 XP 变异体（XP-V）的患者其缺陷基因是编码 DNA 聚合酶 η 的基因，因此，这些患者的嘧啶二聚体损伤不能通过由 DNA 聚合酶 η 催化的、相对错误概率较低的无差错旁路机制处理，而必须通过依赖于包括 DNA 聚合酶 ζ 在内的其他特殊 DNA 聚合酶催化的

易错旁路机制获得处理。易错旁路系统在复制嘧啶二聚体时，引入到子代 DNA 链中的突变不会被切除修复系统所消除。然而，由于这些患者具有正常的切除修复系统，所以只有少数的二聚体损伤需要通过易错系统处理。从而解释了 XP-V 细胞对紫外线照射敏感性相对较低的原因。

聚合酶 η 自身不能执行无差错旁路修复功能，当在子链对应于嘧啶二聚体的位置处插入 2 个 A 后，新合成子链的 3′端不能与 T 形成碱基对，因为模板链上的 T 被封闭在嘧啶二聚体内。由于没有配对的核苷酸的添加，复制型 DNA 聚合酶（α 和 δ）便不能重新起始 DNA 的合成，另外的 DNA 聚合酶可能是聚合酶 ζ 必须执行 DNA 合成的任务。

为什么聚合酶 η 不能简便地继续 DNA 的合成，从而合成足够长的子代 DNA 链，使复制型聚合酶能够重新起始 DNA 的复制？答案在于这是一个易错的过程。尽管“无差错”这一术语用于描述 DNA 聚合酶 η 处理嘧啶二聚体损伤的能力是完全合理的，但该酶在复制正常 DNA 时，是极容易出现错误的。当 Hanaoka、Thomas Kunkel 及同事利用具有裂隙的双链 DNA 分子，在体外检测 DNA 聚合酶 η 的复制保真性时发现，DNA 聚合酶 η 的复制保真性要低于任何已被研究的模板依赖的 DNA 聚合酶：每复制 18～380nt 就会插入一个错误的碱基。相反，DNA 聚合酶 ζ 的准确度要比 DNA 聚合酶 η 高大约 20 倍，因此，对细胞而言，拥有 NER 系统是一件好事，如果没有 NER 系统，DNA 聚合酶 η 只能处理胸腺嘧啶二聚体，而对其他损伤则会束手无策，典型的 XP 患者就是很好的例证。

在处理特定类型的 DNA 损伤时，DNA 聚合酶 η 的跨损伤合成具有特异性。在处理嘧啶二聚体损伤时，DNA 聚合酶 η 能够进行跨损伤合成，但不能在［6-4］光化合物位点处进行跨损伤合成，DNA 聚合酶 η 也能跨越无碱基的 AP 位点。Hanaoka 及同事在体外分析 DNA 聚合酶 α 或 DNA 聚合酶 η 处理不同 DNA 损伤时的跨损伤合成能力，实验所使用的模板具有如下特征：一条链为损伤链，另一条链为 ^{32}P 标记的引物链，其 3′端恰好位于损伤位点的上游。然后向反应体系中加入游离的核苷酸，进行跨损伤合成，最后凝胶电泳检测反应产物。

图 20.44 描述了实验的结果。图 20.44a 表明聚合酶 α 和 η 都能够以完好的 DNA 为模板实现引物的延伸，但是聚合酶 α 在经过 DNA 损伤位点时，不能有效地延伸引物。出现这种现象并不奇怪，因为聚合酶 α 只对正常 DNA 进行准确拷贝，不能处理受损的无信息的 DNA 位点。图 20.44b～图 20.44d 表明聚合酶 η 能够跨越环状嘧啶二聚体（CPD）和 AP 位点进行引物的延伸，但不能跨［6-4］光化合物进行引物延伸。

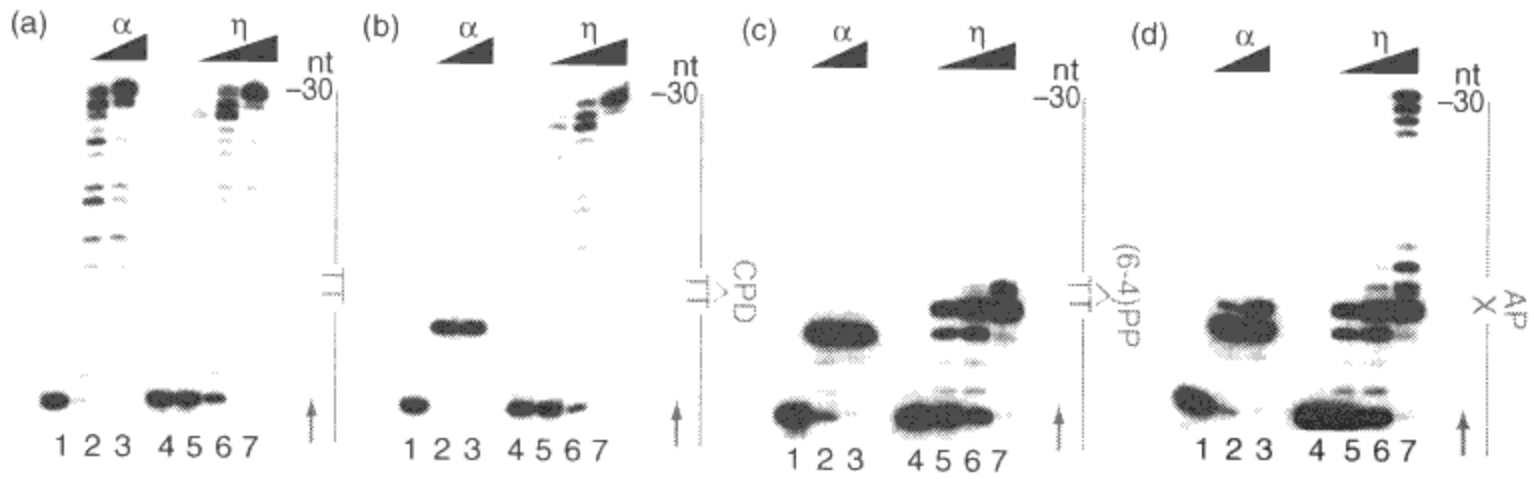


图 20.44 DNA 聚合酶 α 和 η 在未受损和受损模板上的活性。Hanaoka 及同事制备了双链 DNA 分子，其中的模板链为：(a) 损伤；(b) 有一环状嘧啶二聚体（CDP）；(c) 有一个［6-4］光化合物（［6-4］PP）；(d) 一个 AP 位点。非模板链是一段^{32}P标记的引物，通过跨越模板链上的损伤（或正常配对的胸腺嘧啶）而获得延伸。在每一图的边侧标注了这些 DNA 分子的特征。研究者们向反应体系中逐渐增加 DNA 聚合酶 α 或 η 的量，同时加入游离的核苷酸，将反应产物进行聚丙烯酰胺凝胶电泳分析。如果能够成功进行跨损伤合成，那么引物就将被延伸至与模板链等同的长度，即 30nt；否则，DNA 合成在损伤处停止。(*Source*: From Masutani et al., *Cold spring Harbor Symposia* p. 76. © 2000.)

小结　人类细胞具有一种相对无差错的旁路系统，能够在子链对应于嘧啶二聚体的位置处插入 dAMP，从而使胸腺嘧啶二聚体（无胞嘧啶的二聚体）得到准确复制。该系统能够利用 DNA 聚合酶 η 与另一种酶复制越过损伤位点的少数几个碱基。当编码 DNA 聚合酶 η 的基因发生缺陷

时，DNA 聚合酶 ζ 或其他的 DNA 聚合酶代替其执行功能，但这些酶会在子链对应于嘧啶二聚体的位置处随机地插入核苷酸，因此这是一种易错系统。校正 UV 损伤时出现的错误导致 XP 变异体 XP-V 的发现。DNA 聚合酶 η 在处理模板上的胸腺嘧啶二聚体和 AP 位点时，是有活性的，但不能处理模板链上的［6-4］光化合物损伤。该酶并不具有真正的无差错特性。模板上存在缺口时，DNA 聚合酶 η 是已知的模板依赖的聚合酶中复制保真性最差的聚合酶之一。

总结

所有（或绝大多数）DNA 复制均遵循以下几个共同的原则：①双链 DNA 以半保留方式进行复制，分离的亲本链均可作为模板，合成与之互补的一条新链；②*E. coli*（以及其他生物体）的 DNA 复制是半不连续的，一条链是沿复制叉移动的方向连续复制，而另一条链是沿相反方向进行不连续复制，产生 1～2kb 的冈崎片段，从而使两条链都按 5′→3′方向合成 DNA 子链；③起始 DNA 的复制需要引物，*E. coli*细胞内的冈崎片段的合成由长度为 10～12nt 的 RNA 引物引发；④大多数真核生物与细菌的 DNA 是进行双向复制的，ColE1 是 DNA 单向复制的一个事例。

环形 DNA 分子通过滚环机制进行复制。双链 DNA 分子的一条被切割形成一个切口，以另一条完整链为模板使切口的 3′端被延伸，从而将切口的 5′端置换出来。在 λ 噬菌体中，置换链作为后随链进行不连续合成的模板。

作为一种多功能酶，pol I 具有 3 种特定的活性：DNA 聚合酶活性、3′→5′外切核酸酶活性及 5′→3′外切核酸酶活性，酶的大功能域具有前两种活性，而 5′→3′外切核酸酶活性则存在于一独立的小功能域中。在温和蛋白酶水解作用下，大功能域（Klenow 片段）与小功能域分离，形成 2 个蛋白质片段，且仍然具有完整聚合酶所拥有的 3 种活性。Klenow 片段（包括所有已知的 DNA 聚合酶）的结构显示该酶具有一个能够与 DNA 结合的裂隙，Klenow 片段的聚合酶活性位点与 3′→5′外切核酸酶活性位点相距较远。

在*E. coli*细胞的 pol I、pol II 和 pol III 这 3 种 DNA 聚合酶中，只有 pol III 是 DNA 复制所必需的，是细菌 DNA 的复制酶。pol III 核心酶由 α、ε 和 θ 3 个亚基组成。α 亚基具有 DNA 聚合酶活性，ε 亚基具有 3′→5′ 外切核酸酶活性，能够执行校正功能。

DNA 复制的保真性对生命而言是至关重要的，为确保 DNA 复制的保真性，*E. coli* 的 DNA 复制机构具有一个内在的需要引发的校正系统。只有配对的核苷酸才能作为 pol III 全酶起始 DNA 复制的引物。因此，如果偶尔掺入了错误的核苷酸，复制将会出现暂停，直至 pol III 的 3′→5′ 外切核酸酶活性将其切除。而 RNA 引物可成为其降解的标签。

哺乳动物细胞内存在 5 种不同的 DNA 聚合酶。聚合酶 δ 和 α 参与 DNA 两条链的复制，聚合酶 α 能够为两条链合成引物，而聚合酶 δ 催化链的延伸。聚合酶 β 和 ε 似乎是在 DNA 修复中起作用，聚合酶 γ 可能负责线粒体 DNA 的复制。

在复制叉处将双链 DNA 解链的解旋酶是由 *E. coli*的 *dnaB* 基因所编码。细菌单链 DNA 结合蛋白与单链 DNA 的结合力远强于与双链 DNA 的结合力。单链 DNA 结合蛋白以协同方式能够与新形成的单链 DNA 进行紧密结合，从而协助解旋酶发挥作用。覆盖在单链 DNA 上的 SSB 可阻止两条分开的单链重新退火，而且能够发挥保护作用，防止单链 DNA 被降解，此外，SSB 对同源的 DNA 聚合酶还有促进作用。SSB 的这些活性对于细菌 DNA 复制是非常重要的。

当解旋酶对闭合环形 DNA 的两条亲本链解旋时，会在 DNA 分子中引入抵消正超螺旋的张力，必须克服这种张力，否则复制叉的行进将会受阻。旋转机制可以释放这种张力，在*E. coli*中，细菌拓扑异构酶，即 DNA 旋转酶是具有旋转功能的首选者。

乙基甲烷磺酸等烷化剂能够使 DNA 碱基发生烷基化修饰，这种修饰要么直接破坏碱基配对或引起碱基丢失；要么引起 DNA 复制或修复缺陷。

不同种类的辐射会造成不同类型的 DNA 损伤。紫外线的能量相对较低，能够产生嘧啶二聚体等中等程度的 DNA 损伤；而 γ 射线及 X 射线为高能射线，能够使 DNA 周围的分子呈离子化状态，产生能够进攻 DNA 的高活性自由基，引起 DNA 碱基的改变或链的断裂。

因紫外线辐射造成的 DNA 损伤（嘧啶二聚体）可以直接被 DNA 光裂合酶所修复，此酶利用近紫外至蓝光波段的光而将嘧啶二聚体中化学键打开；鸟嘌呤 O6 位点的烷基化可以被自杀性 O6-甲基鸟嘌呤甲基转移酶修复，可以将烷基基团转移到酶自身的一个氨基酸残基上。

碱基切除修复（BER）主要作用于损伤轻微的碱基。由 DNA 糖基化酶起始碱基切除修复过程，该酶能够使碱基对中的损伤碱基外凸，然后切除损伤碱基，形成无嘌呤或无嘧啶位点，在 DNA 修复酶的作用下，切除该位点的磷酸戊糖，由正常的核

苷酸将其取代。在细菌中，由 DNA 聚合酶 I 填补切除修复而缺失的核苷酸；在真核生物中，由 DNA 聚合酶 β 完成填补任务。然而，DNA 聚合酶 β 易出现错误，且自身又无校正活性，因此，由 APE1 执行校正功能。对 DNA 分子内 8-氧代鸟嘌呤位点的修复是 BER 的一个特殊事例，可通过 2 个途径进行修复。由于 oxoG 会与 A 错配，所以，在 DNA 复制结束后由腺嘌呤 DNA 糖基化酶将 A 切除，然而，如果复制没有发生，oxoG 仍然与 C 配对，则由另一种 DNA 糖基化酶，即 oxoG 修复酶将 oxoG 切除。

核苷酸切除修复（NER）主要作用于会导致 DNA 双螺旋发生扭曲的严重损伤。在细菌的核苷酸切除修复过程中，内切核酸酶直接在损伤位点两侧将 DNA 链切断，切除长度包含损伤位点在内的寡核苷酸片段。由 DNA 聚合酶 I 填补空隙、由 DNA 连接酶将切口粘合。

真核生物的核苷酸切除修复通过两个途径完成对损伤的修复，在全基因组 NER（GG-NER）途径中，XPC 与 hHR23 组成的复合体与基因组中的损伤位点结合，导致 DNA 的局部熔解从而起始修复过程。该蛋白质显然能够募集 XPA 和 RPA，随后 TFIIH 加入到复合体中，它的 2 个亚基（XPB 和 XPD）利用自身的 DNA 解旋酶活性使 DNA 的熔解区域得以延伸。RPA 与 2 个切除核酸酶（XPF 和 XPG）结合，将它们定位在被切割链的损伤位点两侧的正确位置上，然后释放出包括损伤位点在内的、长度为 24～32nt 的寡核苷酸片段。转录偶联 NER（TC-NER）与 GG-NER 的机制十分相似，只是由 RNA 聚合酶发挥 XPC 的功能，检测损伤位点并起始 DNA 的熔解，在 TC-NER 与 GG-NER 途径中，一旦损伤的 DNA 被切除，DNA 聚合酶 ε 或 δ 就会填补空隙，由 DNA 连接酶将切口粘合。

通过同源重组或非同源末端连接可修复双链 DNA 断裂损伤。非同源末端连接修复过程需要 Ku 和 DNA-PKcs 的参与，均与 DNA 末端结合，装配成有活性的 DNA-PK 复合体，使 DNA 末端的微同源区彼此排列在一起，一旦微同源区排列在一起，2 个 DNA-PK 复合体彼此进行磷酸化修饰，这种磷酸化作用可以激活 DNA-PKcs 催化亚基的解离并激活 Ku 的 DNA 解旋酶活性，使 DNA 末端发生解旋。末端解旋后，微同源区进行碱基互补配对。最后额外的侧翼 DNA 片段被切除，空隙填补后，DNA 的末端被永久连接在一起。

非同源末端连接和同源重组都需要进行染色体重建。在酵母细胞中，蛋白激酶 Mec1 和 Tel1 被募集到 DSB 位点后，使核小体组蛋白 H2A 的第 129 位丝氨酸残基发生磷酸化修饰，而这种磷酸化作用能够使染色质重建蛋白 INO80 被招募到 DSB 处，该蛋白质利用自身 DNA 解旋酶活性推动核小体远离 DSB 的末端，导致在断裂处形成外突的单链 DNA 3′端，该末端是 NHEJ 和同源重组所必需的。另一种染色质重建蛋白 SWR1 与 INO80 共享多种组分，并也被募集到 DSB 处，可用无磷酸化修饰的 H2A 变异体 Htz1 取代 H2A。使邻近 DSB 位点处已发生磷酸化的 H2A 恢复正常状态。

DNA 复制后遗留的错配错误可以被检测到并获得修复。*E. coli*的错配修复系统通过识别亲本链上 GATC 序列中的甲基化腺嘌呤而将亲本链与子链区分开来，对亲本链的互补链（子链）上的错配进行修复。人类细胞错配修复系统的失效会导致微卫星不稳定性，最终会引发癌症。

细胞可以通过非修复方式来处理 DNA 损伤，方式之一为重组修复，与损伤链互补且具有缺口的 DNA 链与另一个子代双链 DNA 分子中的同源单链进行重组，使缺口问题得以解决，但最初的损伤未获得修复。另一种处理损伤的机制是引发 SOS 应答，该机制即使不可能正确阅读 DNA 损伤区域，也会完成 DNA 复制，导致新合成 DNA 链上出现错误，因此，这种处理 DNA 损伤的机制被称为易错旁路。

人类细胞具有一种相对无差错的旁路系统，能够在子链对应于嘧啶二聚体的位置处插入 dAMP，从而使胸腺嘧啶二聚体（无胞嘧啶的二聚体）得到准确复制。该系统能够利用 DNA 聚合酶 η 与另一种酶复制越过损伤位点的少数几个碱基。当编码 DNA 聚合酶 η 的基因发生缺陷时，DNA 聚合酶 ζ 或其他的 DNA 聚合酶代替其执行功能，但这些酶会在子链对应于嘧啶二聚体的位置处随机地插入核苷酸，因此这是一种易错系统。校正 UV 损伤时出现的错误导致 XP 变异体 XP-V 的发现。

复习题

1. 比较 DNA 复制的全保留复制机制、半保留复制机制及弥散复制机制的差异。

2. 叙述实验并给出结果，证明 DNA 复制方式为半保留复制。

3. 比较 DNA 连续复制、半不连续及不连续复制模式的差异。

4. 叙述实验并给出结果，证明 DNA 复制方式为半不连续复制。

5. 解释在野生型*E. coli*细胞内，标记核苷酸的短时脉冲仅出现在短 DNA 片段中，换句话说，为什么这些细胞 DNA 的复制好像完全是不连续的，而非半不连续。

6. 叙述实验并给出结果，测量冈崎片段上引物的大小。

7. 描述 *B. subtilis* 染色体 DNA 为双向复制，而 colE1 质粒为单向复制的电子显微镜证据。

8. 图示说明 λ 噬菌体 DNA 滚环复制的机制。

9. 图示说明*E. coli* DNA 聚合酶的校正过程。

10. *E. coli* DNA 聚合酶 I 具有哪些活性？在 DNA 复制过程中都具有什么作用？

11. Klenow 片段与完整的*E. coli* DNA 聚合酶 I 有何不同？哪种酶用于缺口翻译？哪种酶用于 DNA 末端补齐？为什么？

12. 对于*E. coli* 的 3 种 DNA 聚合酶，哪个是 DNA 复制必需的，列举相关的证据。

13. pol III 的哪个核心亚基具有 DNA 聚合酶活性？如何确定该亚基？

14. pol III 的哪个核心亚基具有校正活性？如何确定该亚基？

15. 从校正的必要性说明在 DNA 复制过程中引发存在的合理性。

16. 列表说明真核生物 DNA 聚合酶的种类及其功能，并简要列举相关的证据。

17. 比较解旋酶与拓扑异构酶在 DNA 复制过程中的活性异同。

18. 列举 SSB 在 DNA 复制过程中的作用。

19. 解释在超螺旋 DNA 的一条链上产生一个切口即可消除超螺旋的原因。

20. 如何得知 DNA 旋转酶能够使酶上的一个酪氨酸残基与 DNA 分子形成共价键？该共价键形成的作用是什么？

21. 以酵母细胞的 DNA 拓扑异构酶 II 为例，阐述 DNA 的片段穿越模型。

22. 比较 UV 辐射造成的 DNA 损伤与 X 射线或 γ 射线造成的 DNA 损伤之间的差异。

23. 哪两个酶能够催化直接逆转 DNA 损伤的反应？图示说明其作用方式。

24. 比较碱基切除修复与核苷酸切除修复的异同。图示说明这两种修复过程。这两种机制分别负责哪种类型 DNA 损伤的修复？

25. 在人类细胞的碱基切除修复过程中，哪种酶具有校正功能？列举主要的证据。

26. 描述人类细胞 oxoG 修复酶 hOGG1 与 oxoG-C 碱基对形成的复合体及 hOGG1 与正常 G-C 碱基对形成的复合体的晶体结构。这些结构如何阐释 oxoG 可被切除而 G 不能被切除的原因？

27. 转录偶联 NER 与全基因组 NER 有何不同？

28. 简要阐述哺乳动物细胞修复双链 DNA 断裂的非同源末端连接机制，并阐明该过程是如何导致核苷酸在修复位点处发生丢失。

29. 在大多数着色性干皮症中，哪种 DNA 修复系统发生了缺陷？为什么 XP 患者对 UV 光线十分敏感？XP 患者修复损伤的主要替代系统是什么？

30. 在 XP-V 病例中，哪种 DNA 修复系统发生了缺陷？这类患者的皮肤癌的概率低于典型 XP 患者的原因是什么？如果 NER 系统失效，XP-V 患者修复损伤的替代机制是什么？

31. 真核生物细胞修复双链 DNA 断裂损伤需要染色质重建的原因是什么？

32. 图示说明*E. coli*细胞的错配修复机制。

33. 图示说明*E. coli*细胞的重组修复机制。

34. 图例说明*E. coli*细胞的易错旁路系统。

35. 为什么说重组修复和易错旁路修复并不是真正的修复系统？

36. 列举 DNA 聚合酶 η 能够跨胸腺嘧啶二聚体和 AP 位点，而不能跨［6-4］光化合物损伤进行 DNA 复制的原因，以及 DNA 聚合酶 α 均不能跨这些损伤位点进行 DNA 复制的原因。

分析题

1. 为什么不能在自然界观察到 DNA 的完全连续复制？

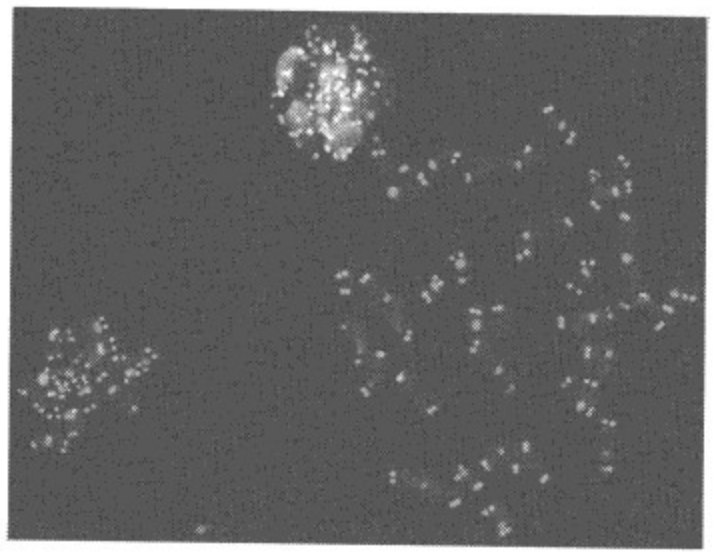
2. 你正在研究一种蛋白质，推测该蛋白质具有 DNA 解旋酶活性，叙述如何通过实验证实这一活性，并给出阳性结果。

3. 你正在研究一种蛋白质，推测该蛋白质具有 DNA 拓扑异构酶活性，叙述如何通过实验证实这一活性，并给出阳性结果。

4. 解释 DNA 损伤与突变之间的差异。以*E. coli* DNA 聚合酶 V 的突变为例，说明 DNA 损伤与突变之间的差异。

翻译　徐启江　校译　李玉花

第21章　DNA复制II：详细机制



人类染色体端粒：绿色为端粒，紫色为着丝粒（*Source: Cal Harley/Geron Corporation & Peter Rabinovitch, Univ. of Washington.*）

通过第20章的学习我们了解到DNA复制是半不连续复制，且引物必须在DNA合成起始前合成。*E. coli* DNA复制需要一些主要蛋白质的参与。DNA复制十分复杂，涉及多种DNA聚合酶。在本章，我们将详细学习*E. coli*和真核生物DNA复制的机制，以及不同生物中DNA复制的3个阶段：起始、延伸和终止。

21.1　复制的起始

我们已经了解到，DNA复制起始于RNA引物的合成，不同生物体合成引物的机制不同，即使侵染*E. coli*的不同噬菌体（大肠杆菌噬菌体）其合成引物的策略也存在很大差异。大肠杆菌噬菌体是探究*E. coli* DNA复制的常用工具，因为它是一类简单生物，主要依靠宿主蛋白来完成自身DNA的复制。

E. coli DNA复制的引发

在第20章已经提到，大肠杆菌噬菌体M13以宿主RNA聚合酶作为引物合成的引发酶（primase，引物合成酶），是偶然发现的有关噬菌体DNA复制引物合成的第一个事例。但*E. coli*及其他噬菌体的引发酶是*E. coli dnaG*基因编码的DnaG蛋白。Arthur Kornberg注意到，*E. coli*及其大多数噬菌体至少还需要一种蛋白质（DnaB，*E. coli dnaB*基因的产物）来完成引物的合成，至少在后随链的引物合成时需要该蛋白质的参与。

Arthur Kornberg及同事在分析单链噬菌体ΦX174的DNA（没有单链结合蛋白）由单链转化为双链时发现了DnaB的重要性，噬菌体第二链DNA的合成需要先合成引物，然后才能进行DNA复制。pol III全酶催化DNA复制，而其他蛋白质则用于引物的合成。在该实验中，Kornberg及同事发现有3种蛋白质是DNA复制所必需的：DnaG（引发酶）、DnaB和pol III全酶。由此可见，DnaG和DnaB显然是引物合成所必需的。Kornberg将参与引物合成的蛋白质复合体称为引发体（primosome），通常是指DnaG和DnaB这两种蛋白质，尽管也有其他蛋白质参与了引发体的组装。

*E. coli*的引发体是可以移动的，但它沿着无SSB结合的环形ΦX174噬菌体DNA移动时，能够不断地重复合成引物。该引发体同样能够在*E. coli* DNA的后随链上引发冈崎片段的合成，而RNA聚合酶或引发酶只能在复制原点处引发DNA的合成。

以噬菌体ΦX174和G4的DNA为模式底物，通过两条基本途径可以鉴定*E. coli* DNA复制体系中的重要组分。第一条途径是运用遗传和生物化学方法，分离DNA复制缺陷突变体，然后用野生型细胞的蛋白质对突变体提取物进行功能互补。突变体提取物只有在加入合适野生型蛋白质后才能在体外进行噬菌体DNA的复制。利用这一分析系统可以有效纯化并鉴别这些蛋白质的特征。第二条途径运用经典的生物化学方法，纯化所有的必需组分，然后在体外重新组装成复制系统。作为重组策略研究者，Kornberg这样总结自己的努力："我一直坚信，只要是细胞能够完成

的事情，一个生物化学家就应该能够完成，他（或她）甚至应该做得更好……换一种说法，利用重组系统能使我们更容易地创造更多的东西。”

***E. coli* DNA 复制起点**　在深入讨论复制引发之前，先研究一下 *E. coli* DNA 复制的独特起始位点 *oriC*。复制起点是正确起始 DNA 复制所必需的 DNA 位点，我们可以通过几种方式定位复制起点，但如何才能确知在起始位点附近起始复制所必需的 DNA 片段长度？方法之一是克隆包含起始位点的 DNA 片段，将其插入到缺失自身复制起点但具有抗生素抗性基因的质粒载体中，然后通过抗生素选择能够进行自主复制的质粒。在抗生素存在条件下能进行复制的细胞一定携带功能性复制起点。一旦获得具有功能性 *oriC* 的质粒，我们就可以通过突变包含 *oriC* 的 DNA 片段，从而发现最短的有效 DNA 序列。*E. coli* 的最小起始位点序列长度为 245bp，在细菌中，复制原点的某些特征是保守的，而且它们的长度也是保守的。

图 21.1 阐明了在 *oriC* 处起始复制的步骤。该复制原点有 4 个 9 聚体序列，其保守序列为 TTATCCACA，两两反向重复。DNA 酶足迹实验表明这些九聚体序列是 DnaA（*DnaA* 基因的产物）的结合位点，因此九聚体序列又被称为 DnaA 盒，DnaA 有助于 DnaB 与复制原点的结合。

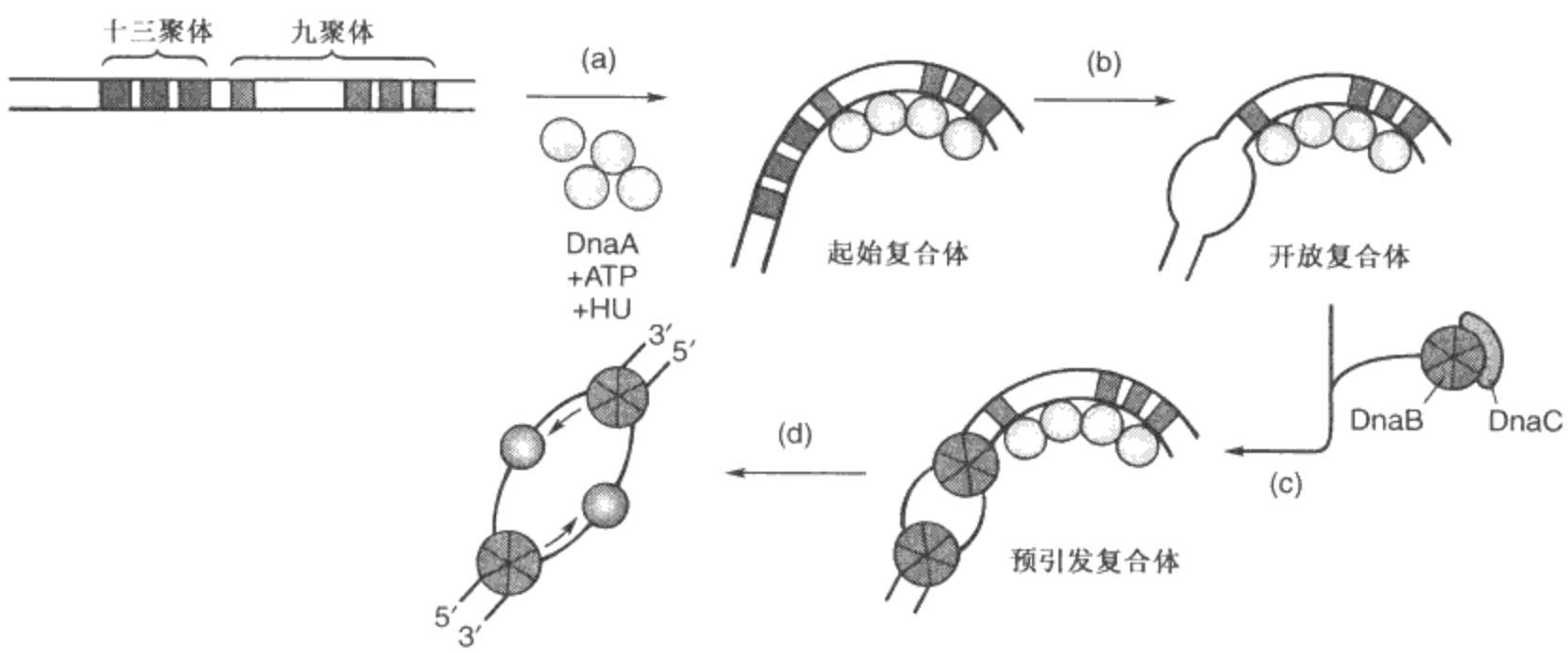


图 21.1　复制在 *oriC* 处引发。(a) 起始复合体的形成。首先，DnaA（黄色）与 ATP 结合形成多聚体，连同 HU 蛋白，DnaA/ATP 复合体结合在包含 4 个九聚体在内的 DNA 序列上，覆盖约 200bp 的 DNA 区域。HU 蛋白可能导致 DNA 序列发生弯曲。(b) 开放复合体的形成。随着 HU 蛋白引起 DNA 序列弯曲，DnaA 的结合显然会破坏邻近的十三聚体重复序列的稳定性，并引起这一区域 DNA 的解链。DnaB 蛋白就会结合在解链的区域。(c) 预引发复合体的形成。DnaC 蛋白结合在 DnaB 蛋白上有助于 DnaB 蛋白与 DNA 结合。(d) 引发。最后，引发酶（紫色）结合到这个预引发体上，形成引发体，引发体合成引物起始 DNA 的复制。箭头代表引物。(*Source*: Adapted from *DNA Replication*, 2/e, (plate 15) by Arthur Kornberg and Tania Baker.)

DnaA 能够促进 *oriC* 左侧十三聚体重复序列的解聚，有利于 DnaB 与复制原点结合而形成开放复合体（open complex）。与开放启动子复合体类似，在开放复合体中，DnaB 结合在解链的 DNA 区域上。在 ΦX 174 引发体组装时，DnaC 与 DnaB 结合并帮助其与复制原点结合。

此证据也强有力地表明，DnaA 能够直接协助 DnaB 与 DNA 的结合。在 R6K 质粒发夹茎环结构的茎部存在一个 *dnaA* 盒，当 DnaA 与之结合时 DnaB（在 DnaC 的帮助下）也能与 *dnaA* 盒结合。由于 *dnaA* 盒没有发生解聚，据此推论 DnaA 会直接影响 DNA 与 DnaB 之间的结合。

至少还有两种其他因子参与了 *oriC* 处开放复合体的形成。第一个因子是 RNA 聚合酶，该酶虽不能像在 M13 噬菌体复制中那样起引发酶作用，但仍发挥着基本功能。因为利福平能够阻断引发体的组装，所以 RNA 聚合酶活性是必需的，其作用可能是合成能形成 R 环（参见第 14 章）的 RNA 片段。R 环可以靠近 *oriC* 位点，但不会插入其中。第二个因子是碱性 DNA 结合蛋白 HU，能够诱导与之结合的双链 DNA 发生弯曲，在 R 环共同作用下，使 DNA 双螺旋失去稳定性，易于解链而形成开放复合体。

最后，DnaB 促进引发酶 DnaG 的结合，完成引发体的组装，从而引发 DNA 的复制。引发体伴随具

有延伸功能的复制体（replisome）而发挥以下作用：首先，重复引发后随链冈崎片段的合成；其次，DnaB作为解旋酶打开DNA双链，为前导链和后随链的复制提供模板。DnaB沿后随链模板按5′→3′方向移动而解旋DNA，复制叉的移动方向与之相同。引发体锚定在后随链模板上引发冈崎片段的合成。

小结 在*E. coli*中，引物的合成需要DNA解旋酶、DnaB和引发酶DnaG在复制原点*oriC*处组装形成引发体。具体组装过程如下：DnaA与*oriC*内的*dnaA*盒结合，与RNA聚合酶、HU蛋白协同作用，使邻近*dnaA*盒最左侧的DNA区域发生解链；随后，DnaB与开放复合体结合，并协助引发酶的结合，从而形成完整的引发体。引发体与复制体一起重复引发后随链冈崎片段的合成。在复制进程中，DnaB发挥解旋酶活性使双链DNA解旋。

真核生物DNA复制的引发

真核生物DNA的复制要比细菌DNA的复制复杂得多，因为真核生物的基因组较大。每条染色体上必须具有多个复制原点，才能解决真核生物复制叉移动较慢的问题，否则，就不可能在较短的时间内（仅为几分钟的S期）完成复制。由于复制原点的多重性以及其他因素，鉴别真核生物DNA复制原点的研究工作要远远滞后于原核生物。但是，当分子生物学家面临这样一个复杂问题时，他们通常将该问题诉诸如病毒等简单系统，从而获得有关宿主的线索。根据这样的策略，科学家在1972年鉴别出了猴病毒SV40的DNA复制原点。下面我们将依次探讨真核生物及酵母DNA的复制原点。

猴病毒SV40的DNA复制原点 由Norman Salzman和Daniel Nathans领导的两个研究小组分别于1972年鉴别出了猴病毒SV40的DNA复制原点，并发现该原点能够起始双向复制。Salzman的研究策略是用*Eco*RI在一特定位点切割正在复制的SV40 DNA分子（尽管*Eco*RI在此之前不久才被发现并确定其性质，但Salzman已知道SV40 DNA上只有一个*Eco*RI的酶切位点），切割后，Salzman及同事在电子显微镜下观察这些分子（图21.2）。研究者只观察到1个复制泡，说明在SV40 DNA分子上只有1个复制原点。而且还观察到复制泡向两个方向生长，在单一复制原点处形成2个相向远离的复制叉。分析研究发现，复制原点与*Eco*RI酶切位点之间的距离长度约为基因组全长的33%，但我们无法分辨哪一个方向是远离*Eco*RI位点的，因为SV40 DNA为环形分子，且图片上除了单一的*Eco*Rl位点外，再无其他任何标记。但是，Nathans用另一种限制性内切核酸酶（*Hind* II）进行实验，并将实验结果与Salzman的实验结果进行综合，最终将复制原点定位在与SV40控制区重叠的位置上，靠近GC盒及具有72bp重复序列的增强子（图21.3）。

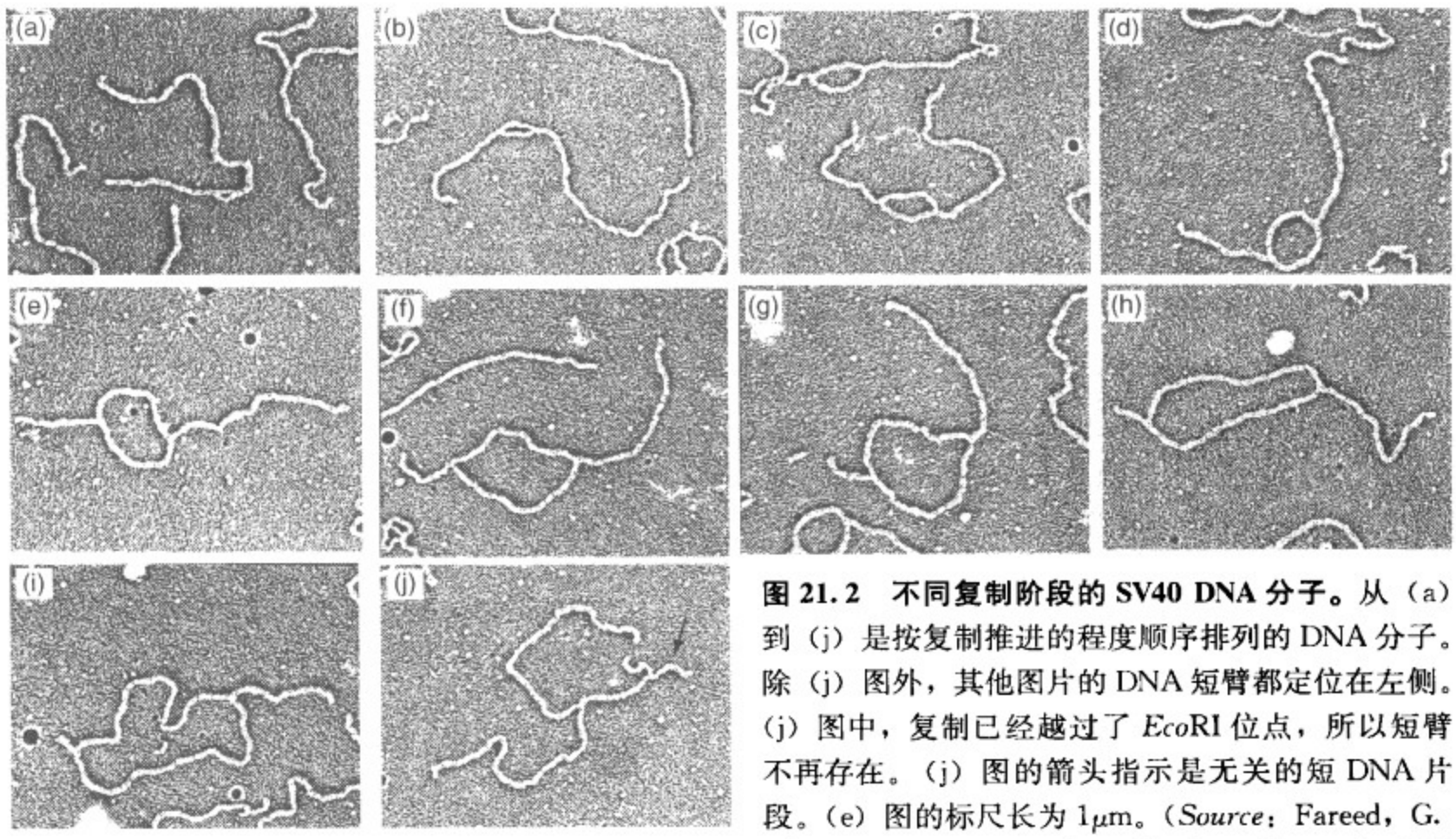


图21.2 不同复制阶段的SV40 DNA分子。从（a）到（j）是按复制推进的程度顺序排列的DNA分子。除（j）图外，其他图片的DNA短臂都定位在左侧。（j）图中，复制已经越过了*Eco*RI位点，所以短臂不再存在。（j）图的箭头指示是无关的短DNA片段。（e）图的标尺长为1μm。（*Source*: Fareed, G. C., C. F. Garon and N. P. Salzman, Origin and direction of Simian virus 40 DNA replication. *Journal of Virology* 10 (1972) f. 4, p. 489, by permission of American Society of Microbiology.）

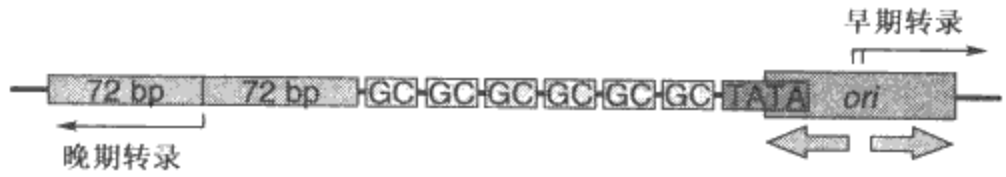


图 21.3　SV40 *ori* 在转录调控区中的定位。ori 核心序列（绿色）包含部分早期区 TATA 盒和早期转录起始位点序列区。粉色箭头表示从起始位点向两侧进行双向复制。黑色箭头表示转录起始位点。

长度为 64bp 的最小 *ori* 序列（*ori* 核心）（图 21.4）包含以下几个重要元件：①4 个 5 聚体（5′-GAGGC-3′），是病毒早期转录区域的主要产物大 T 抗原的结合位点；②15bp 的回文序列，是 DNA 复制起始时的解聚区域；③仅含 AT 碱基对的 17bp 序列，可能有助于回文序列的解旋。

ori 核心周围的一些其他元件也参与了复制的起始。包括 2 个额外的大 T 抗原结合位点（图 21.4）以及位于 *ori* 核心左侧的 GC 盒。GC 盒对复制起始有 10 倍的促进作用，如果缩减 GC 盒数目或被移动至距离 *ori* 核心 180bp 的位置处，其促进复制起始的作用将被减弱或消失。该效应与*E. coli* RNA 聚合酶在 *oriC* 处参与复制起始的效应相类似。不同之处是：SV40 *ori* 不需要转录的发生，转录因子 Sp1 与 GC 结合足以启动复制。

一旦大 T 抗原与 SV40 *ori* 位点结合，即行使解旋酶功能，将 DNA 解旋并为引物的合成做好准备。真核细胞的引发酶与 DNA 聚合酶 α 有关，并且是 SV40 DNA 复制的引发酶。

小结　SV40 复制原点靠近转录控制区域。复制的引发依赖于大 T 抗原与一个 64bp 的 ori 核心及两个邻近位点的结合并行使解旋酶的活性，在 ori 核心区打开一个复制论。引发在引发酶的作用下进行，它与寄主的 DNA 聚合酶 α 有关。

酵母 DNA 的复制原点　目前，有关真核生物 DNA 复制原点的多数信息都是由酵母提供的。酵母作为最简单的真核生物已被广泛地用于遗传学分析，所以，对酵母的遗传学特征已有了很好的了解。早在 1979 年，Hsiao 和 Carbon 研究发现，有一段 DNA 序列能够独立于酵母染色体而自主复制，表明该序列包含有复制原点。将该片段上的 *ARG4*$^+$ 基因克隆到质粒载体中，该载体能够使 *arg4*$^-$ 突变体细胞转化为能够在缺乏精氨酸的培养基上生长的 *ARG4*$^+$ 酵母细胞。任何能够在此培养基上生长的酵母细胞一定整合有质粒中的 *ARG4*$^+$ 基因，且一定以某种方式使基因发生了增殖。通过重组将基因整合到宿主染色体上是基因增殖的途径之一，尽管重组的频率很低，仅为10^{-6}～10^{-7}。Hsiao 和 Carbon 获得了重组频率高达 10^{-4} 的 *ARG4*$^+$ 质粒，在酵母与*E. coli* 之间来回穿梭不会使质粒结构发生任何改变，而与酵母基因组发生重组则会使质粒结构发生显著改变。因此，研究者推断，克隆到质粒上的酵母 DNA 片段可能含有复制原点。同样是在 1979 年，Davis 及同事进行了类似的研究实验，携带酵母 DNA 片段的质粒能够将 *trp*$^-$ 酵母细胞转化为 *TRP*$^+$ 细胞，研究者将这段 850bp 长的酵母片段称为**自主复制序列 1**（autonomously replicating sequence 1，ARS1）。

尽管这些早期的研究很有启发性，但研究者并未能证明 DNA 复制确实是从 ARS 序列开始的。为了证实 ARS1 确实具有复制原点的关键特性，Bonita Brewer 和 Walton Fangman 用双向电泳方法检测携带 ARS1 的质粒复制的起始位点。双向电泳技术的原理是，在凝胶电泳时，环状和分支状 DNA 分子在凝胶中的迁移率要低于相同大小的线性 DNA 分子，尤其是在高压或高琼脂糖浓度条件下差异更为明显。Brewer 和 Fangman 制备了以 ARS1 作为唯一复制原点的质粒，使之在同步化的酵母细胞中复制，然后分离复制中间体（replication intermediate，RI），用限制性内切核酸酶处理，使复制中间体线性化，于低电压和低琼脂糖浓度条件下进行第一维电泳，DNA 分子依据其大小进行初步分离，随后在高电压和高琼脂糖浓度条件下进行第二维电泳，从而延缓环状和分支状 DNA 分子在凝胶中的迁移率。最后用标记的质粒特异 DNA 与凝胶上的 DNA 进行 Southern 斑点杂交，检测电泳结果。

图 21.5 为假定长度为 1kb 的片段在形成各种分支状和环状 RI 分子后在凝胶上电泳行为的理想模式。简单“Y”形（图 21.5a）是 1kb 线性片段的右末端带有一个微小的“Y”形，其电泳行为基本与线性片段相同。随着复制叉由右向左移动，“Y”不断变大，在第二维电泳（垂直）中迁移速率变慢，随着“Y”变得更大，其茎部很短，电泳行为更接近于 2kb 的片段，在图 21.5a 中用带有一垂直短线的水平线条表示这种形状的 DNA 分子。随着复制叉的继续移动，越来越像线性分子，其迁移速率也相应升高，直至复制叉到达片段的末端。在这一点上，这些分子的形状，即迁移率十分接近于 2kb 的线性片段。这种行为导致弧形模式的产生，弧形的顶端

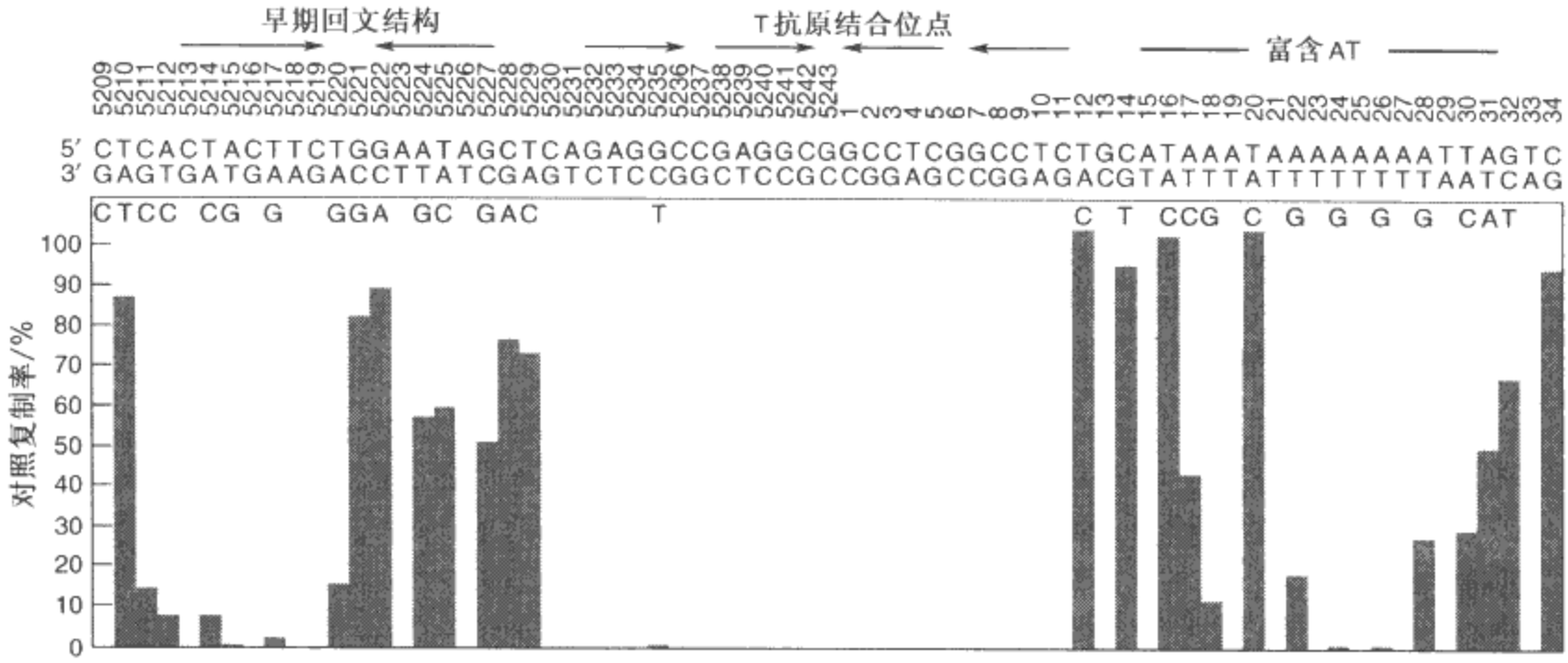


图 21.4 SV40 *ori* 核心区的结构。在碱基对位数下显示 SV40 DNA *ori* 核心区序列，即从 5211 位碱基对到 31 位碱基对间的序列。回文序列区、T 抗原结合位点、A-T 区的位置在顶部标出。箭头指示出 T 抗原结合位点上 4 个五聚体的方向。柱形图显示了在许多位点中进行点突变后的影响，值得注意的是，只有一个片段的位置是突变的。（*Source*：Adapted from Deb，S.，A. L. DeLucia，A. Koff，S. Tsui and P. Tegtmeyer，The adenine-thymine domain of the simian virus 40 core antigen directs DNA bending and coordinately regulates DNA regulation. *Molecular and Cellular Biology* 6：4582，1986.）

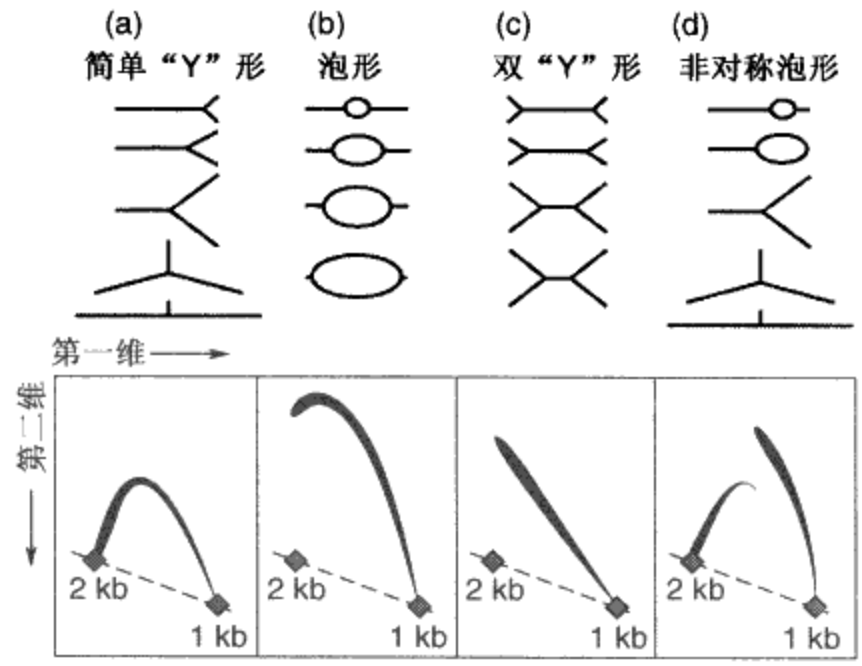


图 21.5 各种复制中间体类型在双向凝胶电泳中的理论上的迁移特征。图（a）～（d）的上半部分显示了正在生长的简单“Y”形、泡形、双“Y”形及非对称泡形，其中非对称泡形在复制过程中可以转化成单“Y”形。图的下半部分描述了推测的各种正在生成的复制中间体在 1kb 到 2kb 大小内与其线性形式（虚线所示）电泳率的偏离。（*Source*：Adapted from Brewer，B. J. and W. L. Fangman，The localiztion of replication origins on ARS plasmids in *S. cerevisiae*. *Cell* 51：464，1987.）

对应于线性片段复制到一半时形成的“Y”形分子，此时片段的形状最不像线性分子。

图 21.5b 为泡状片段预期的电泳行为。同样以 1kb 线性片段为例，只是在片段中部有一微小的泡状结构，随着泡状结构不断变大，迁移速率变得越来越慢，电泳后，形成图底部所示的弧形泳动曲线。图 21.5c 为双“Y”形片段的迁移行为模式，当 2 个复制叉移动到接近片段中部位置时，RI 的分支逐渐增大，其迁移率也相应地呈线性递减。图 21.5d 为泡状结构处于片段非对称位置处时的电泳行为模式，复制开始时为泡状，当复制叉移动越过片段右末端的限制酶酶切位点后，由泡形变为“Y”形。RI 的迁移率反映了这种跃变：迁移曲线开始像是泡形分子的迁移行为，然后突然变成“Y”形分子的迁移行为，明显不连续的迁移曲线准确地显示，当复制叉移动越过限制酶酶切位点后，片段由泡形转变为“Y”形。

这种特性对于确定复制原点的位置特别有用。例如，在图 21.5d 中，曲线的中部发生了间断，在第一维电泳中此迁移率对应于 1.5kb 片段，这说明“Y”形分子的每条臂的长度达到 500bp。假定 2 个复制叉移动速度相同，则复制原点与片段右末端相距 250bp。

现在，让我们了解这项研究工作的实施过程。Brewer 和 Fangman 选用只能对携带 ARS 的质粒切割 1 次的限制性内切核酸酶，如果复制原点确实位于

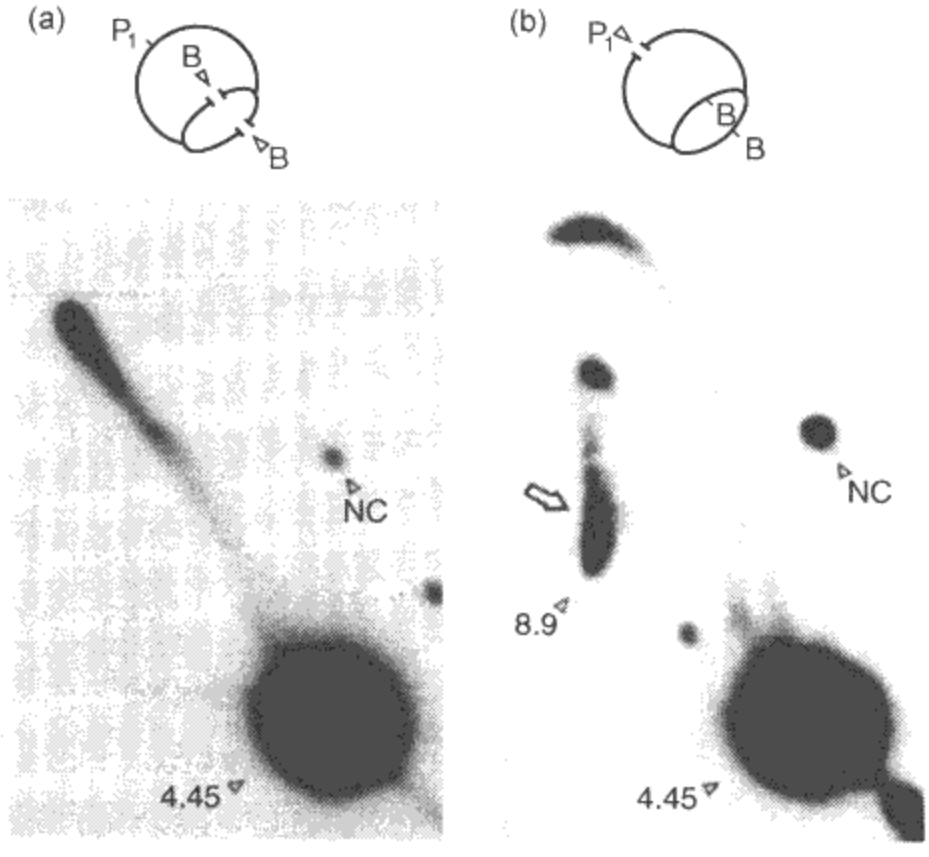


图 21.6　ARS1 复制原点的定位。（a）*Bgl*II 酶切 2μm 质粒后的结果。上图：*Bgl*II 酶切 RI 后的预期示意图，假定起点位于 ARS1 中 *Bgl*II 酶切位点附近。泡状结构含有已经复制的 DNA，因此有两个 *Bgl*II 位点（箭头标为 B），从两个位点切开后产生图中所示的双“Y”形中间体。下图：实验结果显示推测的双“Y”形中间体的直线臂。（b）*Pvu*I 酶切质粒后的结果。上图：*Pvu*I 酶切 RI 后的预期示意图。假定原点位于 ARS1 中 *Pvu*I 位点的对面。下图：实验结果显示生成的弧形结构，其近末端是不连续的。这也正是我们预想有关当复制叉移动越过片段右末端的 *Pvu*I 酶切位点后由泡形变为“Y”形的结果，这两个结果都证实了有关复制原点位于 ARS1 内的推测。NC 表示具有切口的环形分子；较大的空心箭头指向较大的“Y”形或非对称“Y”形分子在复制叉移动越过 *Pvu*I 酶切位点的电泳结果，数字表示 kb 的大小。（*Source*：Brewer，B. J. and W. L. Fangman，The localization of replication origins on ARS plasmids in S. *cerevisiae*. *Cell* 51（6 Nov 1987）f. 8，p. 469. Reprinted by permission of Elsevier Science.）

ARS1 内，那么，切割位点将会提供十分有价值的信息，图 21.6 的上方显示了 2 个酶切位点的位置，图版下方为实验结果。首先值得我们注意的是，放射自显影图像十分简明，与图 21.5 中模式相对应，这说明在 ARS1 内只有 1 个复制原点，否则，会形成由各种不同 RI 组成的混合物，实验结果也会变得更为复杂。

预测的 ARS1 内原点邻近 *Bgl*II 酶切位点（B，图 21.6a），若用 *Bgl*II 酶切 RI，则会产生双“Y”形 RI 分子。正如图 21.6a 下半部分所示，放射自显影图像确实几乎呈线形，与“Y”形 RI 的预测结果一致。图 21.6b 显示，*Pvu*I 酶切位点（P）与预测的复制原点之间的距离几乎为质粒 DNA 长度的一半，因此，经 *Pvu*I 酶切后会产生如图 21.6b 上方所示的泡形 RI 分子，图 21.6b 下方的放射自显影图像显示了 Brewer 和 Fangman 观察到的电泳迁移曲线的不连续性，同预想的那样，当 1 个复制叉移动到 *Pvu*I 酶切位点时，泡形 RI 分子在最末端转化为极不对称的大“Y”形 RI 分子，这种跃变会使电泳迁移曲线发生间断。这 2 个实验结果证实了事先预想，如果 ARS1 包含复制原点，则该复制原点位于 *Bgl*II 酶切位点附近。

York Marahrens 和 Bruce Stillman 通过接头扫描实验确定了 ARS1 内的重要区域。研究者构建了一个类似 Brewer 和 Fangman 使用过的重组质粒，该重组质粒包含：①ARS1 位于一段 185bp 的 DNA 序列内；②一个酵母着丝粒；③选择性标记 *URA*3，能够使 *ura3-52* 酵母细胞在缺乏尿嘧啶的培养基上生长。研究者用 8bp *Xho*I 接头系统置换 ARS1 区域内的正常 DNA 序列，然后进行接头扫描（参见第 10 章）。用接头扫描突变体转化酵母细胞后，在缺乏尿嘧啶的培养基上进行筛选。由于每一个质粒都携带有着丝粒，可保证质粒的正确分离，但一些包含 ARS1 突变序列的转化细胞比包含野生型 ARS1 序列的转化细胞生长要慢。其可能的解释是，由于 ARS1 发生突变而导致 DNA 复制不足。

为验证这个假设，Marahrens 和 Stillman 在含尿嘧啶的非选择培养基上将所有的转化细胞培养 14 代，然后将转化细胞培养在缺乏尿嘧啶的选择培养基上，以观察哪一种转化细胞不能稳定生长传代，其原因可能是这些不稳定质粒携带的突变干扰了 ARS1 的功能。图 21.7 是该实验的结果。ARS1 有 4 个重要的区域，根据其对质粒稳定性影响作用的递减次序而分别被命名为 A、B1、B2 和 B3。元件 A 长 15bp，包含一个长 11bp 的 ARS 保守序列：

$$5'\text{-}\begin{matrix}T\\A\end{matrix}TTTA\begin{matrix}TA\\CG\end{matrix}TTT\begin{matrix}T\\A\end{matrix}\text{-}3'$$

当该序列发生突变时，ARS1 会丧失全部活性。其他区域的突变则不会产生如此强烈的影响，特别是在选择培养基上。然而，通过凝胶电泳分析表明，B3 发生突变会对质粒 DNA 弯曲产生明显影响，柱形图下方的染色凝胶显示，B3 区域发生突变后，质粒 DNA 的电泳迁移率增高，Marahrens 和 Stillman 认为这是 ARS1 发生弯曲的结果。

存在于 ARS1 中的 4 个重要区域是否是 ARS 发

挥功能所必需的？Marahrens 和 Stillman 构建了一个合成的 ARS1，其 4 个区域为野生型序列，ARS1 中 4 个区的间隔距离与野生型情况一致，只是这 4 个区两两间的距离是随机的。一个携带合成 ARS1 序列质粒的酵母细胞在非选择培养基上几乎与携带野生型 ARS1 质粒的酵母细胞一样能够稳定传代，这说明，通过接头扫描实验确定的这 4 个区对 ARS 发挥功能来讲是必需的。研究者用 11bp 的 ARS 保守序列取代长 15bp 的 A 区后，质粒的稳定性会显著降低，这说明 A 区的其他 4bp 对于 ARS 活性也是必需的。

小结 酵母复制原点包含的自主复制序列（ARS）拥有 A、B1、B2 和 B3 4 个重要区域。A 区长 15bp，其中 11bp 序列在 ARS 中是高度保守的。B3 区可在 ARS1 内形成重要的 DNA 弯曲。

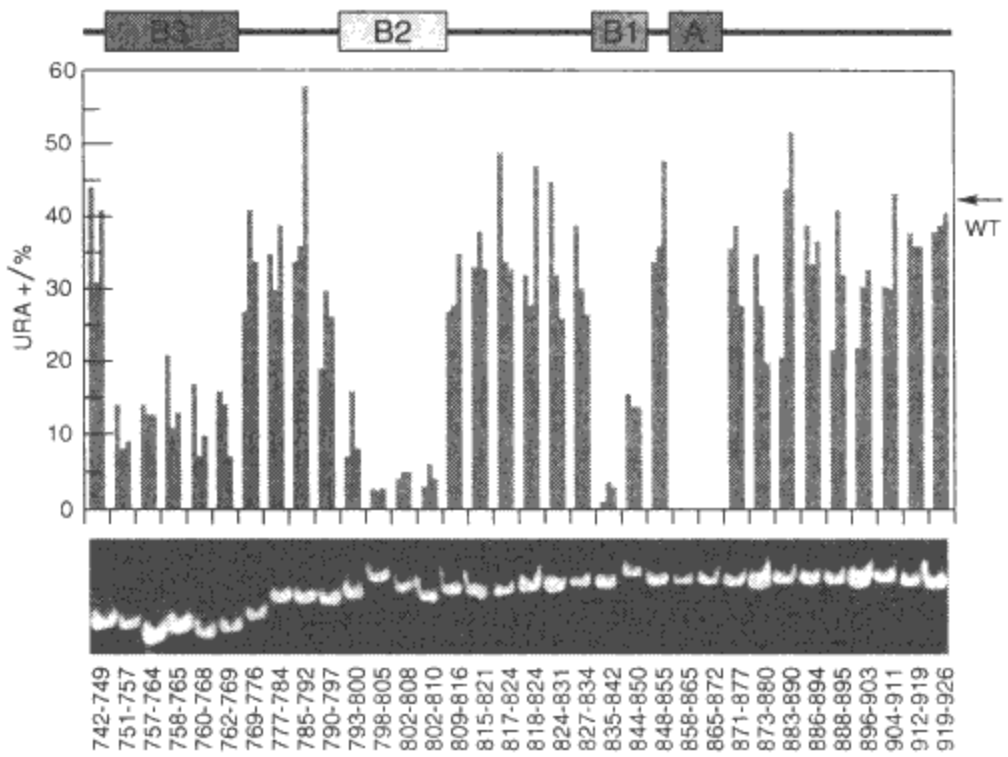


图 21.7 ARS1 接头扫描分析。Marahrens 和 Stillman 用接头替换含酵母着丝粒和 *URA3* 选择性标记质粒中的全部 ARS1 序列。为检测突变体的复制效率，将突变体于基本培养基上培养 14 代，然后于选择培养基（无尿嘧啶）上筛选。相邻 3 个竖条显示 3 个独立实验对每个突变质粒鉴定的结果。结果以含质粒的酵母细胞（分析它们的生长情况）所占的百分比表示。应该注意的是，在基本培养基上，即使是含野生型质粒的酵母细胞也仅有 43%能够稳定传代（右侧箭头所示）。通过实验，鉴定出 4 个重要区域（A、B1、B2 和 B3）。突变区域在图底部用碱基位数标明。底部的染色凝胶显示各突变质粒电泳的迁移率。注意，B3 突变质粒泳动速率的变化说明该质粒 DNA 构象发生了改变。（*Source*：Marahrens，Y. and B. Stillman，A yeast chromosomal origin of DNA replication defined by multiple functional elements. *Science* 255（14 Feb 1992）f. 2，p. 819. Copyright © AAAS.）

21.2 延伸

一旦引物在某处结合，便可起始真正的 DNA 复制（延伸）。pol III 全酶是 *E. coli* 的 DNA 复制酶，DNA 聚合酶 δ 是真核生物的 DNA 复制酶，承担着前导链和后随链的延伸。*E. coli* DNA 复制系统研究得较为透彻，研究数据显示该系统通过一种精巧方式协调后随链和前导链的合成，并保持 pol III 全酶与模板的结合从而使复制具有很高的进行性，以保证较快的复制速度，下面我们将学习 *E. coli* DNA 复制的延伸机制。

复制的速度

Minsen Mok 和 Kenneth Marians 研究了 pol III 全酶在体外进行 DNA 复制时，复制叉的移动速度。研究者构建了能够进行滚环复制的环形模板（图 21.8）。该模板的尾部用 ^{32}P 标记且带有用于引发复制的 3′-自由羟基。Mok 和 Marians 将此模板与全酶、预引发体蛋白和 SSB 温育，或将模板只与 DnaB 解旋酶温育，每隔 10s 提取标记 DNA 产物，通过凝胶电泳测量产物的长度。图 21.9a 和图 21.9b 是这两个反应的实验结果，图 21.9c 为两个反应中复制叉移动速率曲线图，730nt/s 的移动速率接近于体内约 1000nt/s 的移动速率。

此外，延伸过程具有很高的进行性。进行性是指酶长时间催化 DNA 合成而不与模板脱离造成重新起始的性能。由于重新起始是一个耗时的过程且在 DNA 复制过程中几乎没有可被浪费的时间，所有延伸过程必须具有很高的进行性。为测量进行性，

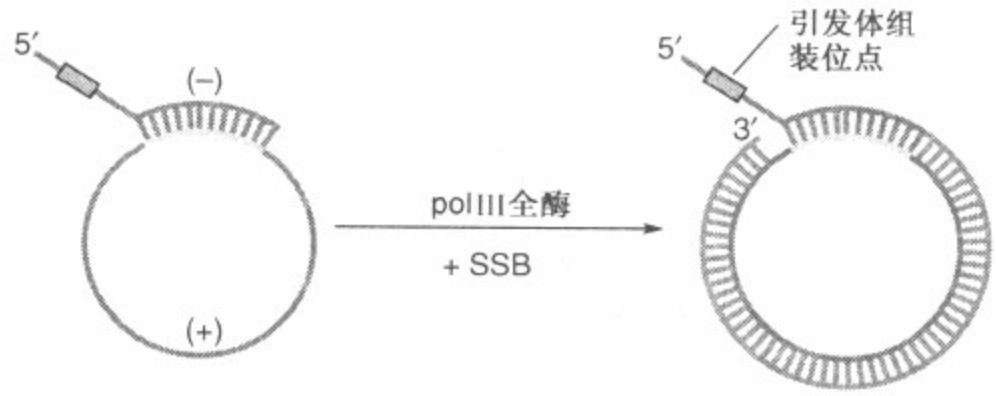


图 21.8　用于体外测定复制叉移动速率的人工合成模板。Mok 和 Marians 将 f1 噬菌体的具 6702nt 的正链（红色）与一段引物（绿色）一起退火，两者杂交的区域为 282nt（黄色）。这个引物含有预引发体装配位点（橙色）。利用 pol III 全酶和单链结合蛋白 SSB 延伸引物而产生负链（蓝色），这一产物是进行多轮滚环复制的双链模板，3′-自由羟基可作为引物。（*Source*：Adapted form Mok，M. and K. J. Marians，The *Escherichia coli* preprimosome and DNA B helicase can form replication forks that move at the same rate. *Journal of Biological Chemistry* 262：16645，1987.）

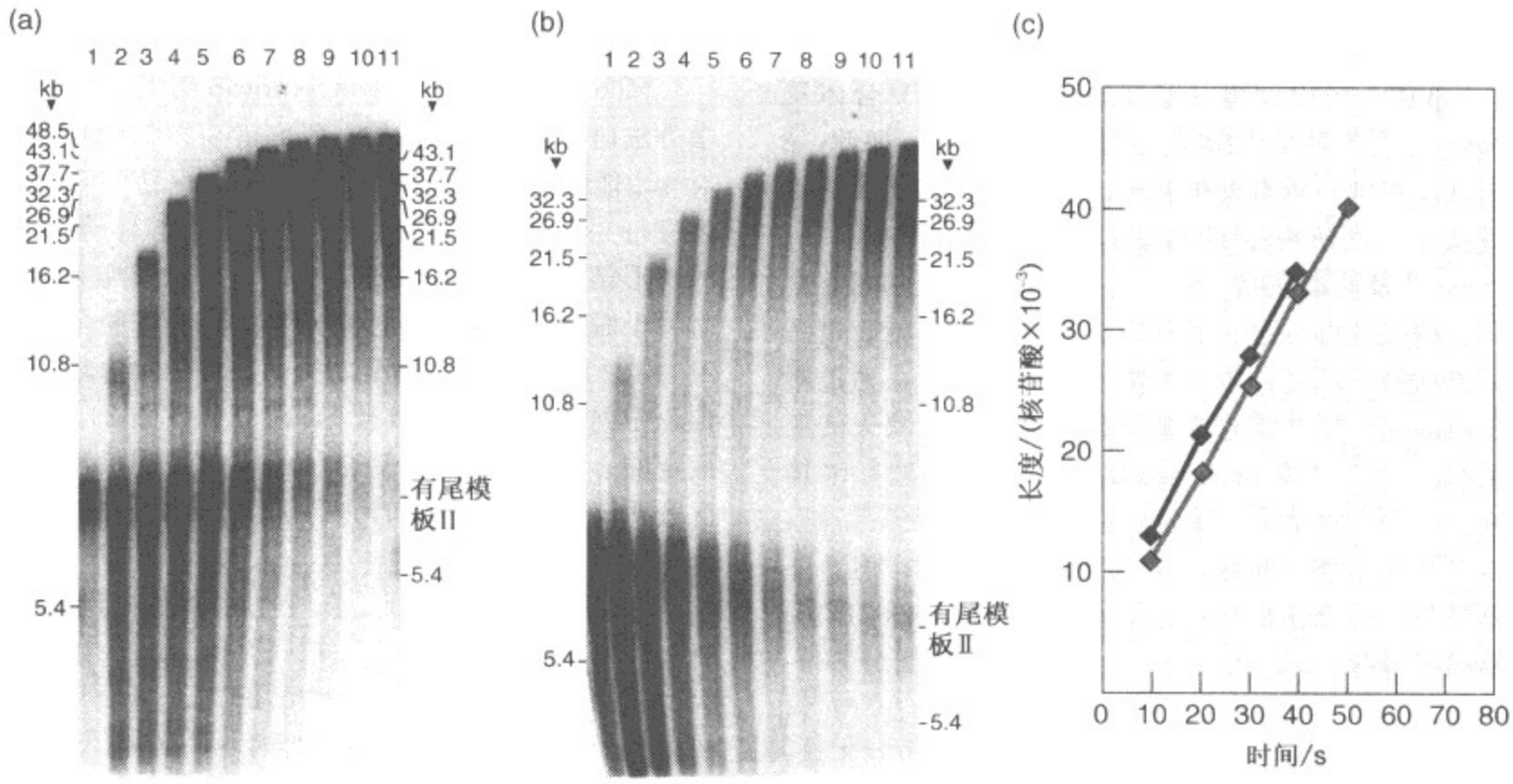


图 21.9　体外测量复制叉的移动速率。Mok 和 Marians 以图 21.8 所示具标记尾部的负链为模板，在体外使之与 pol III 全酶反应，在体系中加入：（a）预引发体蛋白（不包含 DnaG）；或（b）DnaB。每隔 10s 提取标记 DNA 产物，电泳并放射自显影检测。条带 1 表示刚开始即 0s、条带 2 表示反应进行至 10s，电泳迁移率与分子质量成对数关系而不是线性函数。（c）红色线与蓝色线分别表示 a 中的前 5 个时间点和 b 中的前 4 个时间点放射自显影结果。（*Source*：Mok M. and K. J. Marians，The *Escherichia coli* preprimosome and DNA B helicase can form replication forks that move at the same rate. *Journal of Biological Chemistry* 262 no. 34（5 Dec 1987）f. 6a-b，p. 16650. Copyright © American Society for Biochemistry and Molecular Biology.）

Mok 和 Marians 进行了与图 21.9 所示相同的延伸分析实验，只是在反应体系中添加了竞争性 DNA 分子 poly（dA）或全酶 β 亚基的直接抗体，当全酶与模板脱离后，这两种物质可阻止酶重新起始复制。无论竞争剂存在与否，延伸速率没有发生改变，表明全酶在延伸引物全过程中并没有脱离模板，其延伸的长度可达 30kb，无论在体内或体外，全酶均具有很高的进行性。

小结　在体外，pol III 全酶能够以 730nt/s 的速率合成 DNA，慢于体内约 1000nt/s 的合成速率。无论在体内或体外，全酶均具有很高的进行性。

pol III 全酶与复制的进行性

pol III 核心酶自身的聚合酶活性很低，在合成约 10nt 寡核苷酸链后就会脱离模板，然后不得不耗费约 1min 的时间才能再与模板和新生 DNA 结合。

这与细胞内的情况形成了鲜明反差，在细胞内复制叉的移动速度为 1000nt/s。显然，核心酶缺失了某些重要组分。

该组分能赋予全酶的进行性特征，使之在聚合 50 000nt 之前保持与模板的结合，而核心酶在聚合 10nt 之前就会脱离模板。为什么会有如此大的差异？全酶的进行性应归功于**滑动钳**（sliding clamp），滑动钳能够使全酶长时间地结合在模板上。全酶的 β 亚基是滑动钳功能的执行者，但仅靠其自身不能与前起始复合体（核心酶及 DNA 模板）结合，需要**钳装载器**（clamp loader）γ 复合体将滑动钳 β 亚基装载到前起始复合体上。γ 复合体由 γ、δ、δ′、χ 和 ψ 亚基组成。在本节，我们将探讨 β 钳及钳装载器的功能活性。

β 钳　可以设想 β 亚基通过与核心酶复合体及 DNA 结合而使 pol III 核心酶获得进行性。如此，β 钳就能将核心酶系束在 DNA 上并保持与 DNA 的附着状态。在研究这种可能性的过程中，Mike O'Donnell 及同事证明 β 和 α 亚基间存在直接的相互作用。研究者将不同的亚基组分混合，通过凝胶过滤的方法使亚基复合体与游离亚基单体分离，通过凝胶电泳探明复合体的亚基组分，通过添加缺失的亚基和测量 DNA 的合成分析复合体活性。图 21.10 描述了这些实验结果，作为核心酶的组成部分，α 和 ε 亚基彼此结合，进而 α、ε 和 β 组成复合体，但 β 亚基是与 α 结合还是与 ε 结合？图 21.10d 和图 21.10e 给出了答案：β 亚基只与 α 亚基结合（在组分 60～64 出现 2 个亚基高峰），但不会与 ε 亚基结合（在组分 68～67 出现 β 亚基峰值，而在组分76～78 出现 ε 亚基峰值），所以，α 是与 β 结合的核心亚基。

该方案需要在 α 和 ε 复制 DNA 时，β 能够沿着 DNA 滑动这也意味着 β 钳可以保持与环形 DNA 的结合，但可以从线性 DNA 的末端滑落。为验证这种可能性，O'Donnell 及同事开展了图 21.11 所示的实验研究，实验的基本策略是，在 γ 复合体的协助下，将 ^{3}H 标记的 β 二聚体装载到环形双链噬菌体 DNA 上，然后用不同的方式处理 DNA，以观察 β 二聚体是否会与 DNA 解聚，通过凝胶过滤方法分析 β 与 DNA 的结合情况。游离的 β 二聚体从凝胶过滤柱滤出的时间要迟于与 DNA 结合成复合体时滤出的时间。

在图 21.11a 中，用 *Sma*I 处理使 DNA 线性化，然后检测分析 β 钳是否会从 DNA 上滑落。结果表明，β 钳可以继续与环形 DNA 结合，但会在线性 DNA 的末端滑脱而与 DNA 分离。图 21.11b 表明，环形 DNA 分子上的缺口不是 β 二聚体保持与 DNA 结合的原因，因为当 DNA 连接酶消除缺口后，β 二聚体仍然结合在 DNA 分子上。插图为 DNA 连接酶确实消除了缺口提供了电泳证据，因为电泳结果显示，随着缺口分子消失，闭合环形 DNA 分子的量随之增加。图 21.11c 表明，在装载反应体系中添加更多的 β 亚基后，结合在环形 DNA 分子上的 β 二聚体数目随之增多。事实上，每个环形 DNA 分子可装载 20 多分子的 β 亚基。这正是我们预期的许多全酶串联协同复制 DNA 的情形。

如果 β 二聚体会从线性 DNA 末端滑脱，那么在 DNA 末端结合其他蛋白质就应该能够阻止 β 二聚体的滑脱。O'Donnell 研究小组通过实验将两种不同的蛋白质结合在 DNA 末端后，不再出现 β 二聚体从线性 DNA 上滑脱的情况。实际上，即使没有蛋白质连接，DNA 末端的单链尾部也可以充当障碍物阻止 β 二聚体的滑脱。

Mike O'Donnell 和 John Kuriyan 用 X 射线晶体学方法研究了 β 钳的结构。研究者获得的图片可以对 β 钳能够与环形而不是线性 DNA 保持结合的原因进行完美的解释。β 二聚体形成能够环绕 DNA 的环形结构，如同套在细绳上的环，若细绳为线形时易于脱落，但若细绳为环形时就不易脱落。图 21.12 是他们构建的模型之一，β 二聚体为环形结构，中间为成比例的 B 型 DNA 分子。图 21.13 是假定的 β 二聚体/DNA 复合体的空间填充模型，再次表明了 β 二聚体是如何巧妙地环绕住双链 DNA 分子的。

在第 20 章我们提到，真核生物具有进行性因子 PCNA，与细菌 β 钳的功能相同。两者的一级结构无明显的相似性，且真核生物的 PCNA 蛋白只有原核生物蛋白质的 2/3 大小。然而，Kuriyan 等的 X 射线晶体学方法研究表明，酵母 PCNA 蛋白形成的三聚体在结构上与 β 钳二聚体十分相似，是一个能够环绕 DNA 分子的圆环（图 21.14）。

小结　pol III 核心（αε 或 αεθ）自身没有进行性功能，在脱离模板之前只能复制一短的 DNA 片段。相比之下，核心酶与 β 亚基结合后会以 1000nt/s 的合成速率进行 DNA 复制。β 亚基二聚体形成环形结构，环绕 DNA 模板，并与 α 亚基相互作用，将模板和聚合酶结合在一起，使聚合酶能够长时间地停留在模板上而具有很高的进行性。真核生物的进行性因子 PCNA 蛋白形成环形三聚体结构，通过环绕 DNA 分子而将聚合酶结合在模板上。

钳装载器　O'Donnell 及同事在图 21.15 所示的实验中证实了钳装载器的功能。由于 θ 亚基在体外实验中并不是必需因子，所以研究者以 α 和 ε 亚基取代核心酶，以与引物退火的单链 M13 噬菌体 DNA 为模板。此外，研究者已知高进行性的全酶可在

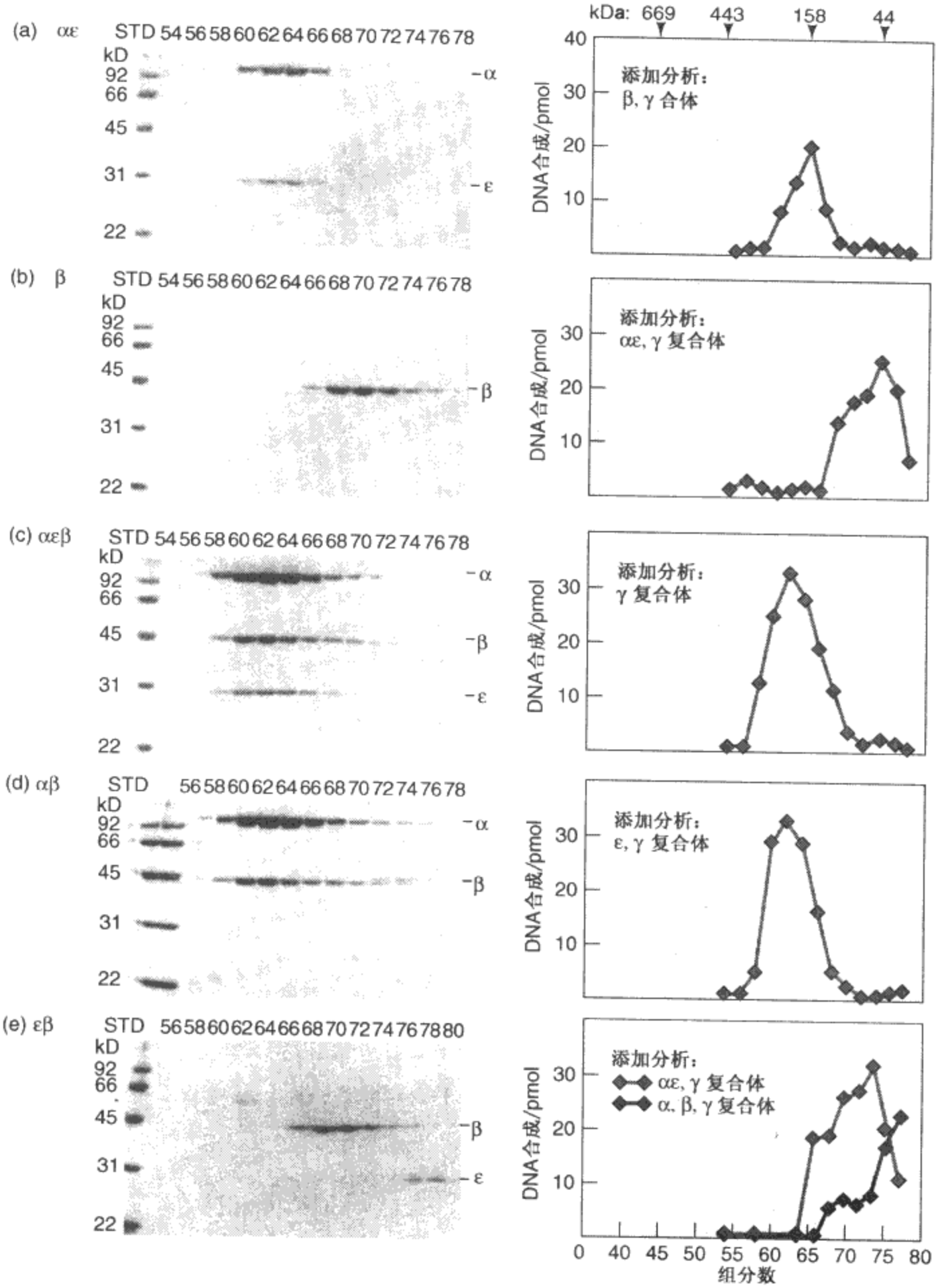


图 21.10　pol III α 和 β 亚基间的相互结合。O'Donnell 及同事将不同的 pol III 亚基混合：（a）α + ε；（b）β；（c）α + ε + β；（d）α + β；（e）ε +β。通过凝胶过滤把混合物中的复合体从游离的亚基中分离出来，将过滤后的组分进行凝胶过滤柱电泳，检测复合体，左边图显示了电泳结果。如果有复合体形成，那么复合体中的亚基就应该出现在同一组分中，如图（a）中的 α 和 ε。右边为在相同的凝胶过滤组分中加入缺失的 pol III 全酶中的亚基，通过分析 DNA 合成以证实复合体的存在。如图（a）中的左侧图，复合体中的 α 和 ε 从凝胶过滤柱中被洗脱下来，所以把其他的基本因子 β 和 γ 加入聚合酶中进行分析。注意：α、ε 和 β 可形成复合体，α 和 β 也可以，但是 ε 和 β 不能。（*Source*: Stukenberg, P. T., P. S. Studwell-Vaughn and M. O'Donnell, Mechanism of the sliding β-clamp of DNA polymerase III holoenzyme. *Journal of Biological Chemistry* 266 no 17 (15 June 1991) figs. 2a-e, 3, pp. 11330-31. American Society for Biochemistry and Molecular Biology.)

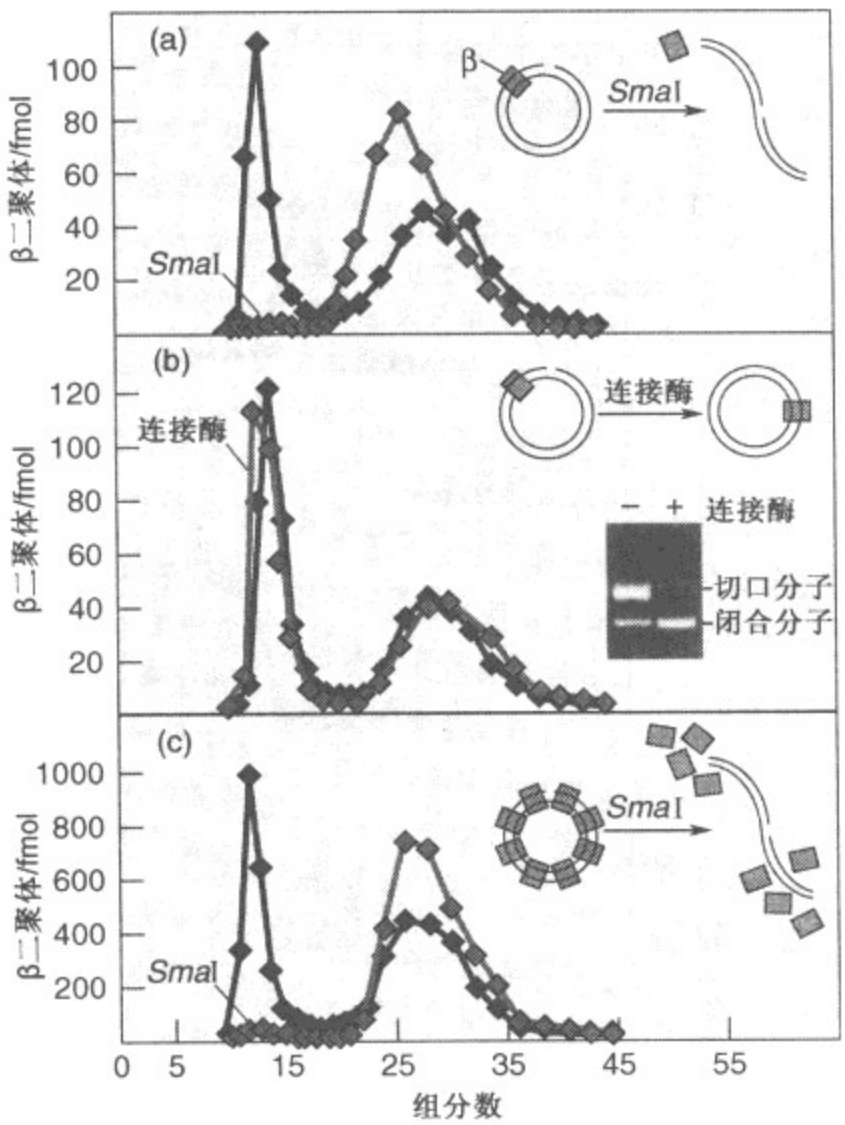


图 21.11　β 钳可从线性 DNA 的末端滑脱。O'Donnell及同事将^{3}H 标记的β二聚体在γ复合物协助下装载到各种 DNA 上，然后按描述的不同方法处理复合体，最后将混合物进行凝胶过滤，将蛋白质-DNA 复合体（大分子从柱子中洗脱得快，约为组分 15）与游离蛋白（相对较小，洗脱得慢，约为组分 28）分离。(a) *Sma*I 线性化 DNA 的效果。DNA 只被 *Sma*I 酶切一次后即进行分析（红色）。未切割的 DNA 也做同样分析（蓝色）。(b) 消除模板上缺口的效果。分析前模板上的缺口由连接酶去除（红色），或保留（蓝色）。贴图为连接酶作用 DNA 前后的电泳结果。(c) 许多β二聚体能装载到 DNA 上，在 DNA 线性化后解离。β二聚体加载到 DNA 模板上的比率随β亚基浓度的升高和 DNA 模板浓度的下降而增加。然后，于分析前对 DNA 用 *Sma*I 酶切（红色）或不酶切（蓝色）。(*Source*: Stukenberg, P. T., P. S. Studwell-Vaughn and M. O'Donnell, Mechanism of the sliding β-clamp of DNA polymerase III holoenzyme. *Journal of Biological Chemistry* 266 no. 17 (15 June 1991) fig. 3, p. 11331. American Society for Biochemistry and Molecular Biology.)

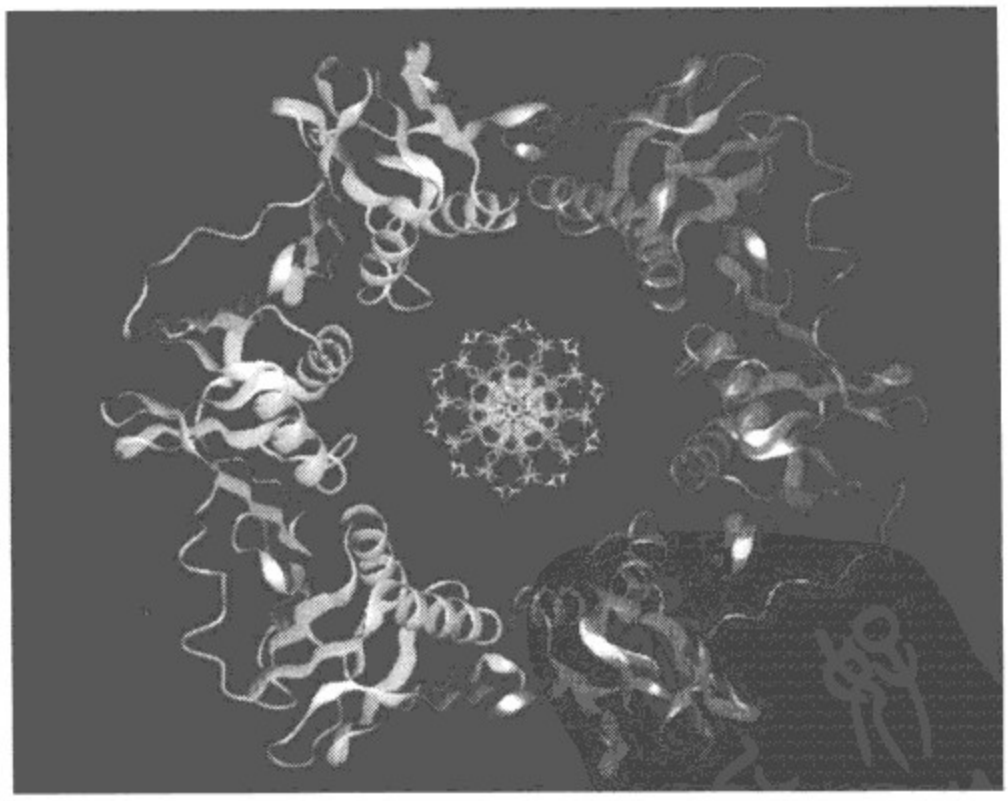


图 21.12　β二聚体/DNA 复合物模型。β二聚体用带状表示，其中α螺旋呈螺旋管状，β折叠呈扁平带状。β单体分别用黄色和红色表示。从横切面看，DNA 模型位于β二聚体形成环中间的假定位置处。(*Source*: Kong, X.-P., R. Onrust, M. O'Donnell and J. Kuriyan, Three-dimensional structure of the beta subunit of *E. coli* DNA polymerase III holoenzyme: A sliding DNA clamp. *Cell* 69 (1 May 1992) f. 1, p. 426. Reprinted by permission of Elsevier Science.)

15s 时间内完成该模板 DNA 的复制，但还不确知 αε 核心在相同的时间内复制 DNA 的量。因此，他们推论：在 20s 的脉冲时间内，所有进行性聚合酶分子均有机会完成一轮复制，这样，复制出的 DNA 轮数应与进行性酶分子数相同。图 21.15a 表明，在 αε 核心及β亚基存在的条件下，每飞摩尔（fmol 或 10^{-15}mol）的γ复合体会完成10fmol DNA 分子的复制，所以，γ复合体具有催化功能：每分子γ复合体能够保证多分子进行性聚合酶的形成。插图为复制产物凝胶电泳结果，同预期的进行性复制结果一致，DNA 产物均为全长分子。

该实验结果表明γ复合体自身不能提供进行性。但能发挥催化功能，使某种物质添加到核心聚合酶上而使之具有进行性。由于β是该实验中唯一的另一

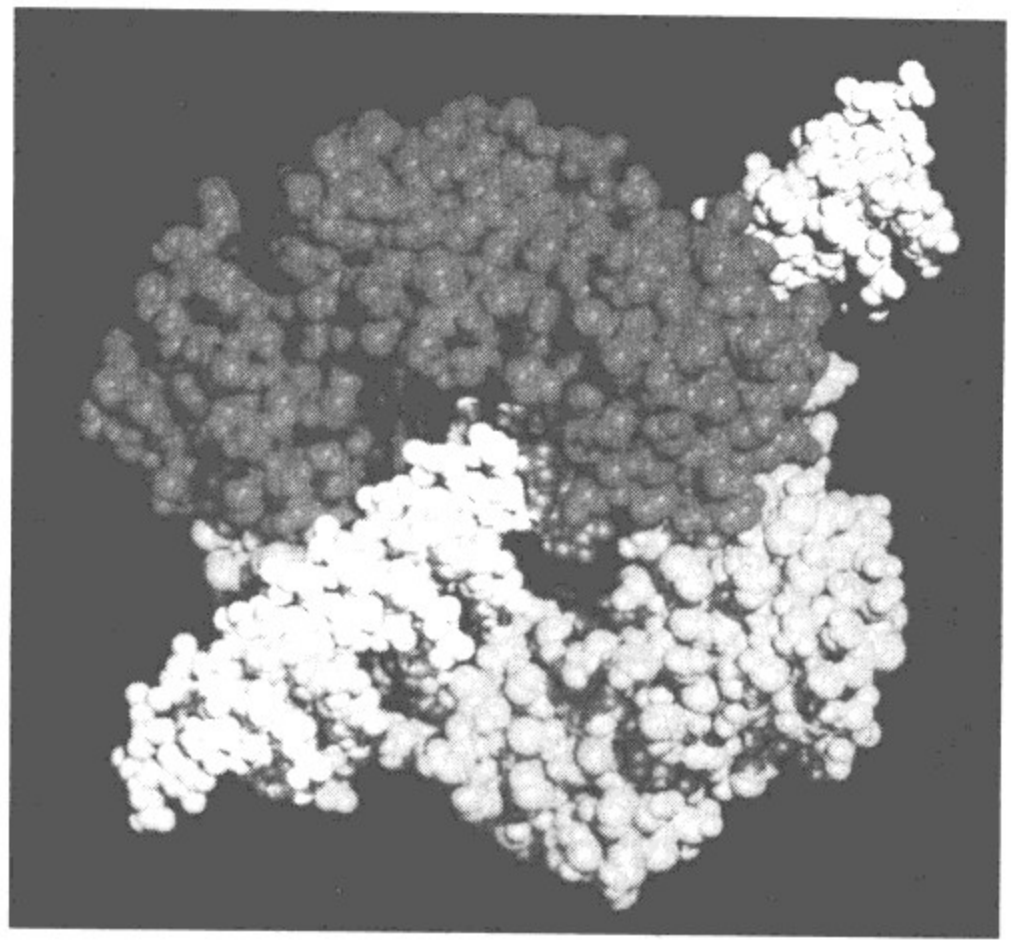


图 21.13　β二聚体/DNA 复合物的空间填充模型。β单体分别用黄色和红色表示，DNA 用绿色和白色表示。DNA 的位置是假设的，但可能性很大。（*Source*：Kong，X.-P.，R. Onrust，M. O' Donnell and J. Kuriyan，Three-dimensional structure of the β subunit of *E. coli* DNA polymerase III holoenzyme：A sliding DNA clamp. *Cell* 69（1 May 1992）f. 3，p. 427. Reprinted by permission of Elsevier Science.）

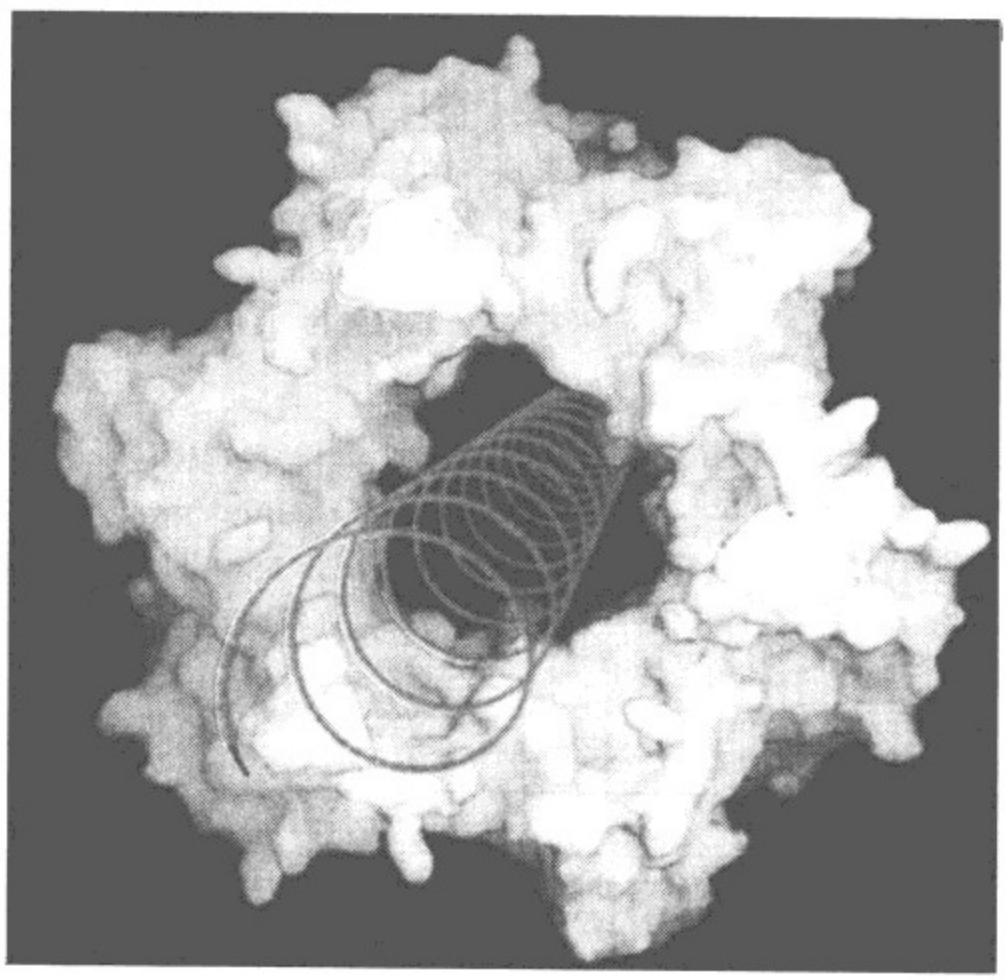


图 21.14　PCNA-DNA 复合物模型。PCNA 三聚体各单体用不同的色彩表示。三聚体的形状是通过 X 射线晶体衍射分析出来的。红色螺旋代表与 PCNA 三聚体结合的 DNA 糖-磷酸骨架可能的位置。（*Source*：Krishna，T. S. R.，X.-P. Kong，S. Gary，P. M. Burgers and J. Kuriyan，Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell* 79（30 Dec 1994）f. 3b，p. 1236. Reprinted by permission of Elsevier Science.）

聚合酶亚基，它很可能是进行性的决定因子。为求证这点，O'Donnell 及同事将 DNA 模板、^{3}H 标记的β亚基以及未标记的γ复合体混合形成前起始复合体，通过凝胶过滤将复合体从游离的蛋白质中分离出来，然后向各组分中添加 αε，通过分析标记双链环形 DNA（RFII，绿色）的形成来检测前起始复合体。图 21.15b 表明，只有痕量的γ复合体（蓝色）与 DNA 的结合，而大多数标记的β亚基（红色）都保持与 DNA 的结合（用抗γ抗体进行 Western 斑点杂交以检测γ复合体，见图 21.15 的底部）。最值得注意的是，即使γ复合体不再维持与 DNA 的结合，它仍然通过将β亚基装载到 DNA 上而在酶的进行性特性上发挥关键作用。

在实验中，O'Donnell 及同事也估算了前起始复合体中β亚基的化学计量。通过对比β亚基与复合体的飞摩尔数发现，二者的比值为 2.8，接近于

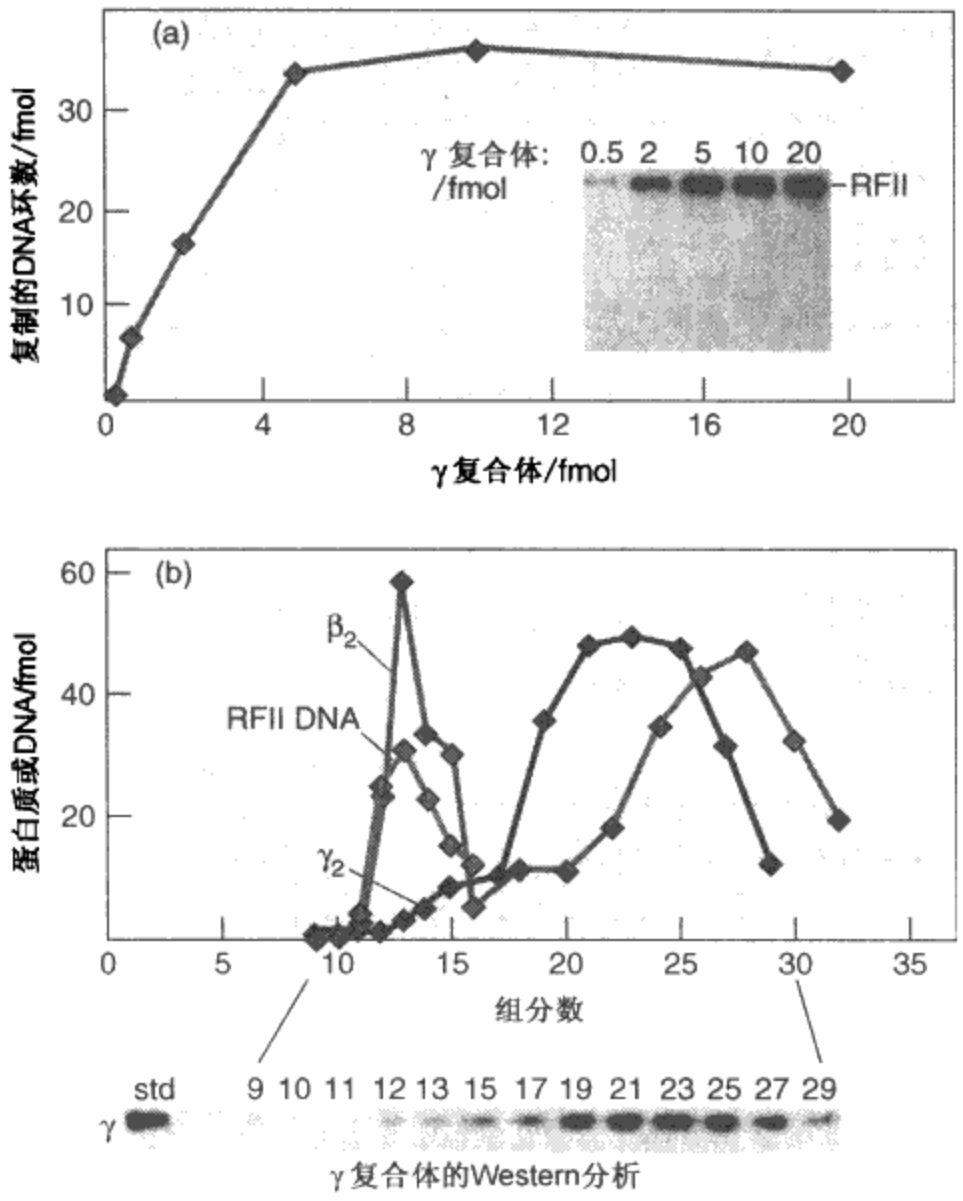


图 21.15　β和γ复合物在进行性中的关联。（a）γ复合体在形成进行性聚合酶过程中发挥催化功能。O'Donnell 及同事向结合有引物、SSB、αε 核心的 M13 噬菌体模板以及 pol III 全酶 β 亚基体系中逐渐增加 γ 复合体（x 轴表示）。然后，加入［α-^{32}P］ATP，20s 脉冲标记 DNA 产物。各取部分产物检测每个反应的放射性，并将数值转化成已复制的 DNA 飞摩尔量。每个反应的另一部分产物进行凝胶电泳，检测全轮复制情况。插图电泳结果：每份产物中的大部分成分是全长的环状 DNA（RFII）。（b）是 β 亚基而不是 γ 复合体与前起始复合体中的 DNA 连接。O'Donnell 及同事在结合有引物并被 SSB 覆盖的模板中加入^{3}H 标记的 β 亚基和未经标记的 γ 复合物，同时加入 ATP，形成前起始复合体。然后，将混合体通过凝胶过滤，使前起始复合体与游离蛋白分离。通过放射性检测每个组分中的 β 亚基，以 γ 抗体为探针进行 Western 斑点杂交，检测 γ 复合体。图中 β 亚基（二聚体）结合到前起始复合物的 DNA 上，但 γ 复合体不与 DNA 结合。（*Source*: Stukenberg, P. T., P. S. Studwell-Vaughn and M. O'Donnell, Mechanism of the sliding [beta] -clamp of DNA polymerase III holoenzyme. *Journal of Biological Chemistry* 266 (15 June 1991) f. 1 a&c, p. 11329. American Society for Biochemistry and Molecular Biology.)

每个复合体结合一个 β 二聚体，与其他研究中 β 亚基以二聚体形式起作用的结果相一致。

在上述讨论中暗含着这样一个事实，即 ATP 是 β 钳装载到模板上的必需条件。Peter Burgers 和 Kornberg 通过分析不需要 dATP 的复制实验证明，ATP（或 dATP）确实是 β 钳装载到模板上的必需因子。分析用的模板为 poly（dA），由 oligo（dT）引发复制，表 21.1 显示，ATP 是 dATP oligo（dT）引物高活性延伸所必需的，尤其是在盐浓度相对较高的条件下。

表 21.1　在 poly(dA)·oligo(dT12)上全酶活性的激活

实验	添加核酸	SSB	[NaCl]eq/(mmol/L)	DNA 合成/(pmol/min)
1	无	−	28	1.0
	ATP 或 dATP	−	28	10
	无	+	34	9.4
	ATP 或 dATP	+	34	33
2	无	−	110	<0.1
	ATP 或 dATP	−	110	8.2
	无	+	110	<0.1
	ATP 或 dATP	+	110	37
	ATPγS	+	110	<0.1

续表

实验	添加核酸	SSB	[NaCl]eq/(mmol/L)	DNA 合成/(pmol/min)
	AMPPNP, d AMPPNP	+	110	<0.1
	GTP, CTP, UTP, dGTP, dCTP, dTTP	+	110	<0.1
	ADP	+	110	<0.1

Source: Burgers, P. M. J. and A. Kornberg, ATP activation of DNA polymerase III Holoenzyme of *Escherichia coli*. *Journal of Biological Chemistry* 257:11469, 1982. Copyright © 1982 The American Society for Biochemistry & Molecular Biology, Bethesda, MD. Reprinted by permission.

钳装载器是如何撬开 β 二聚体使之钳住 DNA 的？O'Donnell、Kuriyan 及同事断定 2 个复合体的晶体结构可为钳的装载提供线索，其中之一就是钳装载器（$\gamma_2\delta\delta'$复合体）活性部分的结构。另一个就是修饰的 β-δ 复合体结构，该复合体由失去二聚化功能的突变体 β（β_{mt}）单体和能够与 β 发生相互作用的 δ 片段组成。

修饰的 β-δ 复合体的晶体结构显示，δ 与 β 单体间的相互作用会通过两种方式在一个界面处削弱 2 个 β 单体间结合。第一，δ 发挥分子扳手的功能，使 β 二聚体的界面构象发生改变，β 亚基也就失去二聚化功能。第二，δ 能够改变其中一个 β 亚基的弯曲度，使之不能与另一个 β 亚基形成正常的环形结构。图 21.16 阐述了这些理念，δ 亚基只与 β 钳中的一个单体结合（在 pol III 全酶中，每个 β 二聚体只结合一个 δ 亚基），所以只能削弱二聚体中的一个分界面，迫使环形结构打开。如果 δ 亚基与 2 个 β 单体都能结合，可能会导致 2 个单体的完全解离。

对复合体结构及我们将在本章后面学习的早期生物化学研究表明，游离的 δ 亚基易与 β 单体结合，但当 δ 与其他亚基形成 $\gamma_2\delta\delta'$装载器后，必须在 ATP 存在条件下才能与 β 结合。ATP 可使装载器构型发生改变，暴露出 δ 亚基使之能够与 β 亚基结合，从而撬开 β 钳。

小结 β 亚基向 DNA 模板上装载需要 γ 复合体（γ、δ、δ′、χ 和 ψ）的协助，γ 复合体在进行性 αδβ 复合体形成过程中起催化作用，在复制过程中，γ 复合体并不与 αδβ 复合体保持结合。钳的装载是一个依赖 ATP 的过程。ATP 释放出的能量改变了钳装载器的构象，使 δ 亚基能够与滑动钳的一个 β 亚基结合，这种结合导致滑动钳开，从而使滑动钳能够环绕 DNA。

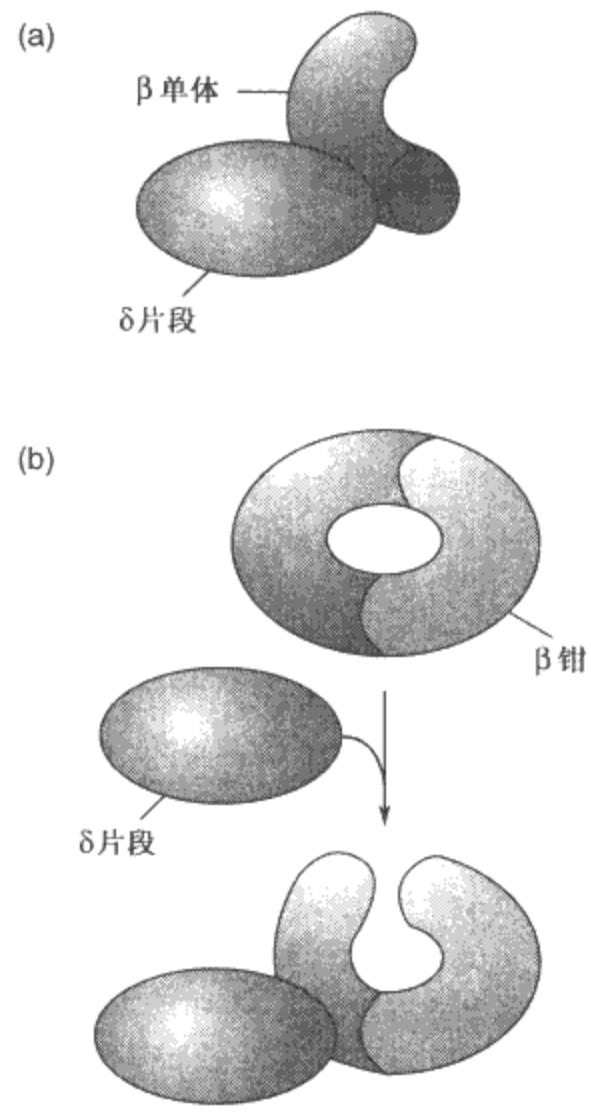


图 21.16　δ 的结合对 β 二聚体影响的模型。（a）δ 片段和 β_{mt}单体形成的复合物的形状。（b）δ 的结合对 β 钳的影响。δ 亚基（或 δ 片段）使 β 二聚体上方交界面之间的结合弱化，并改变了左侧 β 单体的弯曲度，使之不再与另一个单体形成完整的环形结构，从而将 β 钳打开。（*Source*: Adapted from Ellison, V. and B. Stillman, OPening of the clamp: An intimate view of an ATP-driven biological machine. *Cell* 106 [2001] p. 657, f. 3.）

后随链的合成　对 pol III*（无 β 钳的全酶）结构研究发现，该酶包含有 2 个核心酶，通过 τ 二聚体连接到钳装载器上（图 21.17）。通过推论表明：τ 亚基充当核心酶二聚化的物质，α 亚基在天然状态下以单体形式存在，而 τ 亚基以二聚体形式存在。而且，τ 直接与 α 亚基结合，所以结合在 τ 二聚体上的 2 个 α 亚基也会发生二聚化。依次，通过与 2 个 α 亚基的结合，ε 亚基形成二聚体，θ 又通过与 ε 亚基结合而二聚化。全酶包含有 2 个核心酶的事实与 DNA 的两条链都需要复制的事实完全吻合，全酶在沿复制叉移动时，每个核心酶负责一条链的复制，图 21.18 描述了 2 条子链同时合成的模式。

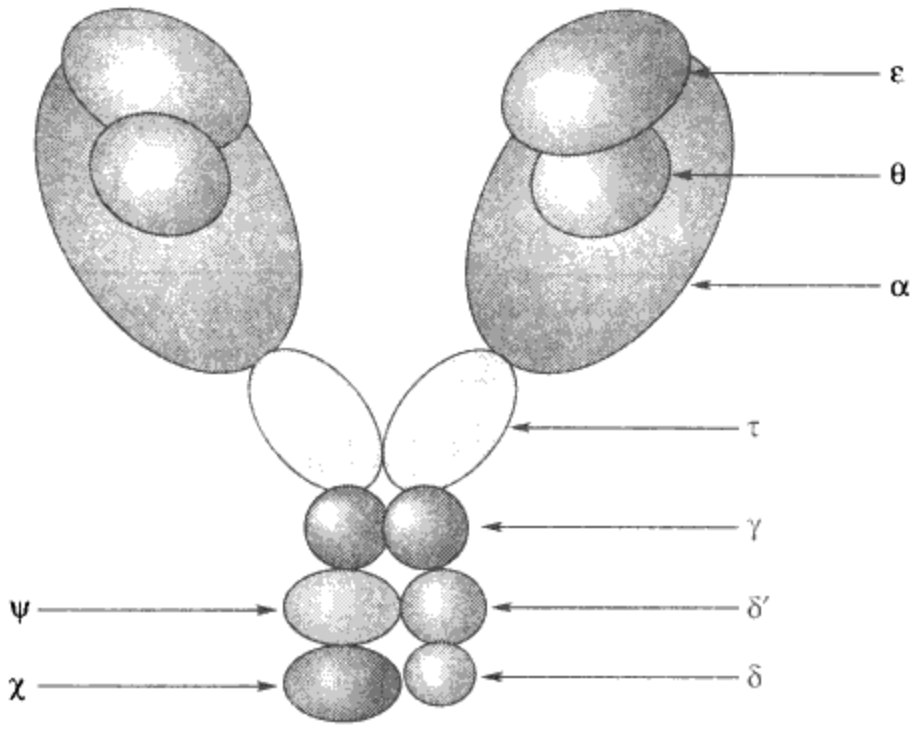


图 21.17　pol III* 亚组合体模型。图中有两个核心与两个 τ 亚基，但只有一个 γ 复合体（γ_2、δ、δ′、χ 和 ψ）。（*Source*: Adapted from Herendeen, D. R. and T. J. Kelly, DNA polymerase III: Running rings around the fork. *Cell* 84: 6, 1996.）

由于后随链的合成是不连续的，这必然涉及核心酶与模板的反复解离与再结合。该模型由此引出2个重要的问题：第一，不连续复制的后随链如何能够与连续复制的前导链保持同步？如果每完成后随链上1个冈崎片段的合成，pol III 核心酶就与模板完全解离，则需要再花费很长的时间才能与模板重新结合，这样就会被前导链远远地落在后面。第二，pol III 核心酶与模板反复不断地进行解离和再结合，又如何与 DNA 复制的高度进行性特征相一致？毕竟 β 钳是复制进行性的必需因子，但一旦钳合在 DNA 模板上，核心酶又如何能够在完成每一个长为 1～2kb 的冈崎片段后与 DNA 解离，然后向前跃进开始下一个冈崎片段的合成？

第一问题的答案似乎是，pol III 核心酶在合成后随链时，并没有真正与模板完全解离，而是通过与合成前导链的核心酶结合而保持与 DNA 的结合，该核心酶通过释放滑动钳而避免自己远离 DNA 模板，从而使核心酶能够发现下一个引物，并在 1s 时间内与模板重新结合，若与 DNA 完全解离，则需要花费较长时间才能重新结合。

要回答第二个问题，我们需要认真考虑 β 钳与钳装载器及核心聚合酶间相互作用的方式。我们将会了解到这两个蛋白质会竞争 β 钳上的同一个结合位点，而 β 钳对钳装载器和核心聚合酶的相对亲和力会发生往复性变化，从而导致核心聚合酶与 DNA 间发生反复性地解离与再结合。我们还将会了解到，钳装载器还可充当卸载器，促进了核心聚合酶与 DNA 间解离与再结合的循环过程。

理论推测过程如下：合成后随链的 pol III* 在完成一个冈崎片段合成后必须与 β 钳脱离，然后再与另一个 β 钳重新结合以便开始下一个冈崎片段合成，（图 21.19）。但是，pol III* 与 β 钳的解离是否真的发生了？为验证这一事实，O'Donnell 及同事将 β 钳和 pol III* 装载到结合有引物的 M13 噬菌体 DNA 模板（M13mp18）上，然后加入 2 条带有多引物的噬菌体 DNA 模板，一条装载有 β 钳（M13Gori），而另一条无 β 钳装载（ΦX174）。然后研究者在适宜的复制条件下进行长时间孵育，从而使原初模板和第二模板都能被复制。研究者知道，DNA 模板 M13mp18 肯定会被复制，但令人感兴趣的问题是：后加的第二模板中哪一条会被复制？是装载有 β 钳的那条还是无 β 钳装载的那条？图 21.20（泳道 1～4）显示，装载有 β 钳的模板 M13Gori 优先被复制。如果将 β 钳装载到另一条模板上后，情况又将怎样呢？泳道 5～8 显示，如果装载有 β 钳，则另一条模板 ϕX174 就会被优先复制。如果 pol III* 与原有的钳保持结合，它应该能够复制任何一条第二模板，而不管第二模板是否装载有 β 钳。所以，该实验结果暗示，pol III* 确实与模板、β 钳发生了解离，如果装载有另一个 β 钳，则 pol III* 就会与其他模板（或同一模板的其他部位）结合。

为验证该结论，研究者用［γ-^{32}P］ATP 对 β 钳进行磷酸化修饰而使 β 钳被 ^{32}P 标记，用 ^{3}H 标记 θ 亚基，或 τ 亚基，或 γ 复合体而使 pol III* 被标记。然后研究者进行如下实验：在只有 dGTP 和 dCTP 存在的条件下，使这些标记的复合体闲置在有缺口的模板上；或在 4 种 dNTP 都存在的条件下，使标记的复合体能够填补模板上的缺口，直至终止。最后，将反应混合物进行凝胶过滤，分析判断 2 种标记物是否发生了分离。当聚合酶闲置在模板上时，标记的 β 钳和 pol III* 在模板上保持结合状态；相反，当发生复制终止时，pol III* 与留在 DNA 模板上的 β 钳解离。O'Donnell 及同事观察到，不管标记 pol III* 中的哪一个亚基，都出现相同的情况。所以

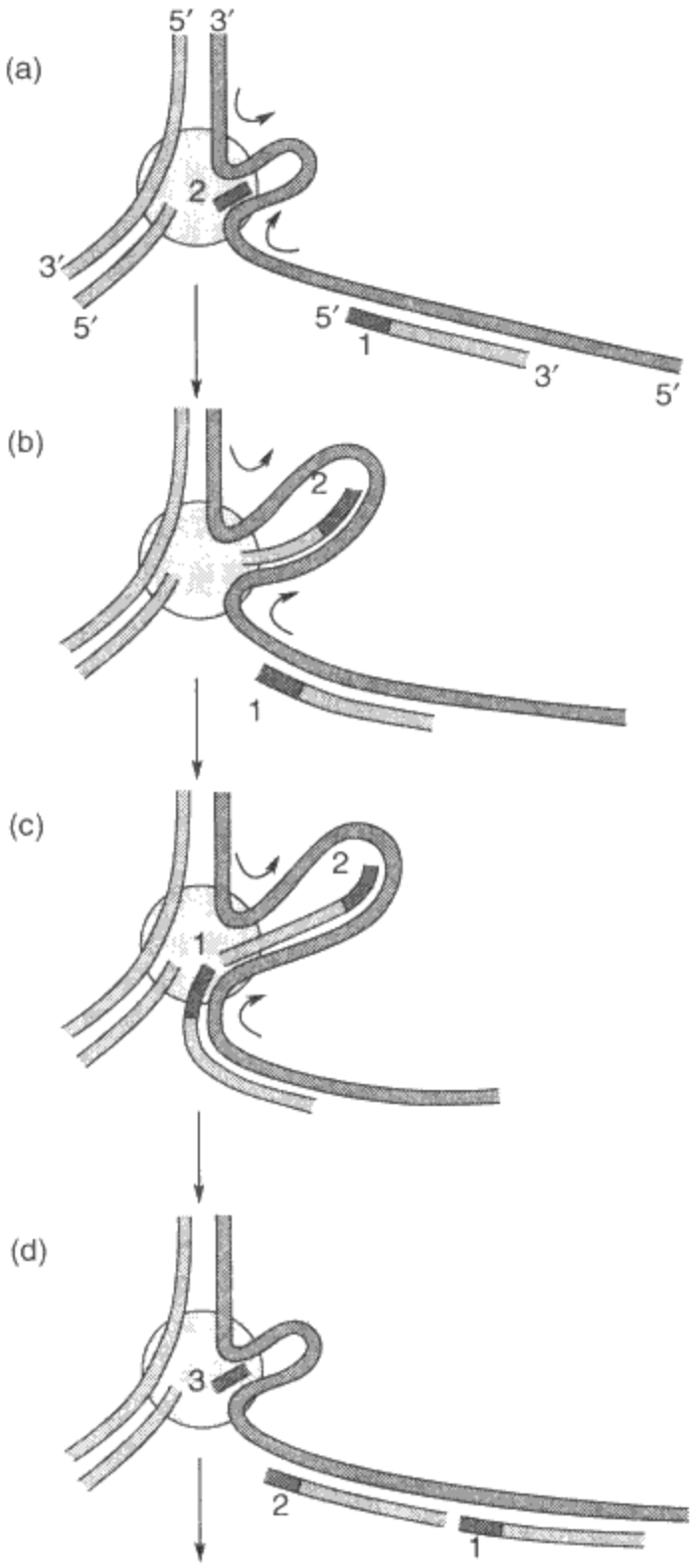


图 21.18　DNA 两条链同时合成的模型。（a）后随链（蓝色）穿过引发体（金色）形成一个环，由引发酶合成一段标记为 2 的新引物（红色）。可以看到已经合成的冈崎片段（绿色，其红色引物标记为 1）。前导链及其新合成的子链在左侧（灰色），图中并没有考虑前导链的延伸。（b）后随链模板由上至下穿过复制体而形成较大的环，如箭头所示。环底部的运动（下部箭头）使第二个冈崎片段得以延伸。（c）第二个冈崎片段的进一步延伸使其末端靠近第一个冈崎片段引物。（d）复制体释放环状结构，使引发酶能够合成新的引物（标记为 3）。至此上述过程又将重新开始。

在复制终止时，是全酶而不只是核心酶与 β 钳和 DNA 模板发生解离。

E. coli 基因组的大小为 4.6Mb，后随链复制时

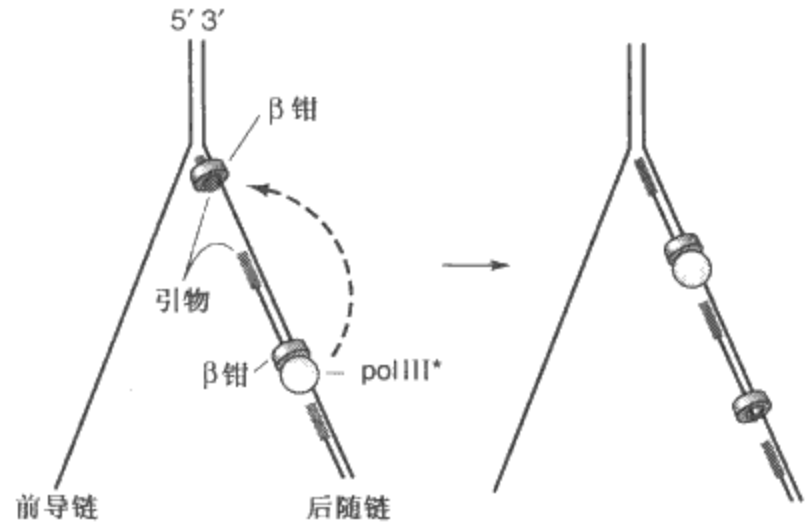


图 21.19　后随链上的 pol III* 复合体与在 β 钳间的循环模型。当 pol III* 复合体（黄色）完成一个冈崎片段的合成后，在已合成的冈崎片段的引物前遇到缺口。这意味着 pol III* 复合体需要从 β 钳（蓝色）上解离下来移动（虚线箭头）到下一个装载到引物（红色）上 β 钳处，起始下一个冈崎片段的合成。（*Source*：Adapted from Stukenberg, P. T., J. Turner and M. O'Donnell, An explanation for lagging strand replication: Polymerase hopping among DNA sliding clamps. *Cell* 78: 878, 1994.）

的冈崎片段仅为 1～2kb，这意味着在每个模板上将有 2000 多次引发事件的发生，至少需要 2000 个 β 钳的参与。由于*E. coli* 细胞内只有约 300 个 β 二聚体，如果不能循环利用，β 钳将很快被消耗殆尽，因此需要 β 二聚体能够与 DNA 模板解离，但这种解离会发生吗？为查明事实，O'Donnell 及同事将多个 β 钳组装在缺口模板上，通过凝胶过滤除去所有的其他蛋白质，然后添加 pol III* 并再次进行凝胶过滤。图 21.21a 充分表明，当存在 pol III* 时，β 钳会与模板解离；但当不存在 pol III* 时，β 钳则不会与模板解离。图 21.21b 证实，这些被释放的 β 钳能够被再次装载到受体模板上。

通过前面的学习，我们已经很清楚地了解到，β 钳既能与核心聚合酶也能与 γ 复合体（钳装载器）发生相互作用，在 DNA 合成过程中，β 钳必须与核心酶结合从而保持聚合酶与模板的结合，随后，β 钳又必须与模板解离，从而可移动到新的模板位点，并与另一个核心聚合酶结合，起始新冈崎片段的合成。当然，向新 DNA 位点的移动需要 β 钳再次与钳装载器发生相互作用。但仍然存在一个关键问题：细胞是如何协调 β 钳与核心聚合酶和钳装载器之间发生的结合转换？

要回答这个问题，首先应了解核心聚合酶、钳装载器在何时，以怎样的方式与 β 钳发生相互作用。O'Donnell 及同事首先回答了作用方式的问

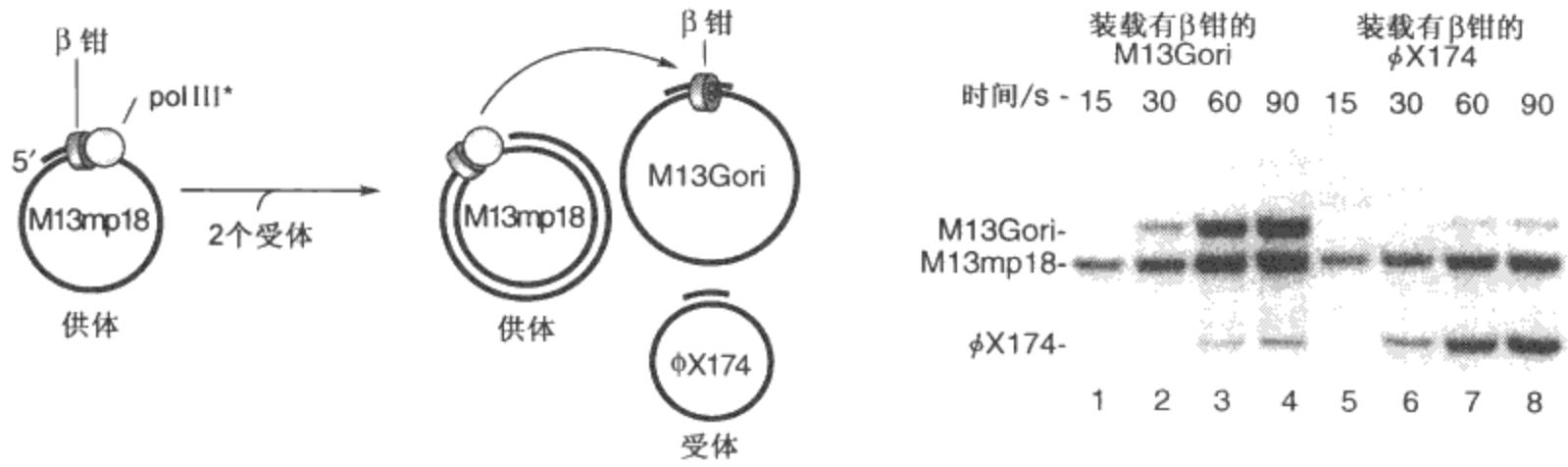


图 21.20　循环模型的检测。如果将 pol III* 复合体与 β 钳组装到带有引物的模板（M13mp18，左上）上然后提供 2 个带有引物的受体模板，其中一个模板装载有 β 钳（M13Gori），另一个模板无 β 钳装载（ϕX174），在它完成原初模板的复制后，pol III* 复合体会选择装载有 β 钳的模板（M13Gor）继续进行复制。O'Donnell 及同事开展了这项实验，在足够长的时间内确保供体模板和受体模板均能得以复制。研究者加入标记的核苷酸从而使复制产物得以标记，然后用电泳检测复制的标记 DNA 产物。DNA 复制产物的电泳结果（底部）显示，装载有 β 钳的受体模板被复制。一方面，当 β 钳装载在 M13Gori 受体模板上时，该模板优先被复制；另一方面，当 β 钳装载在 ϕX174 模板上，则此模板优先被复制。复制的模板位置标记在左侧。（*Source*：Stukenberg，P. T.，J. Turner and M. O'Donnell，An explanation for lagging strand replication：Polymerase hopping among DNA sliding clamps. *Cell* 78（9 Sept 1994）f. 2，p. 878. Reprinted by permission of Elsevier Science.）

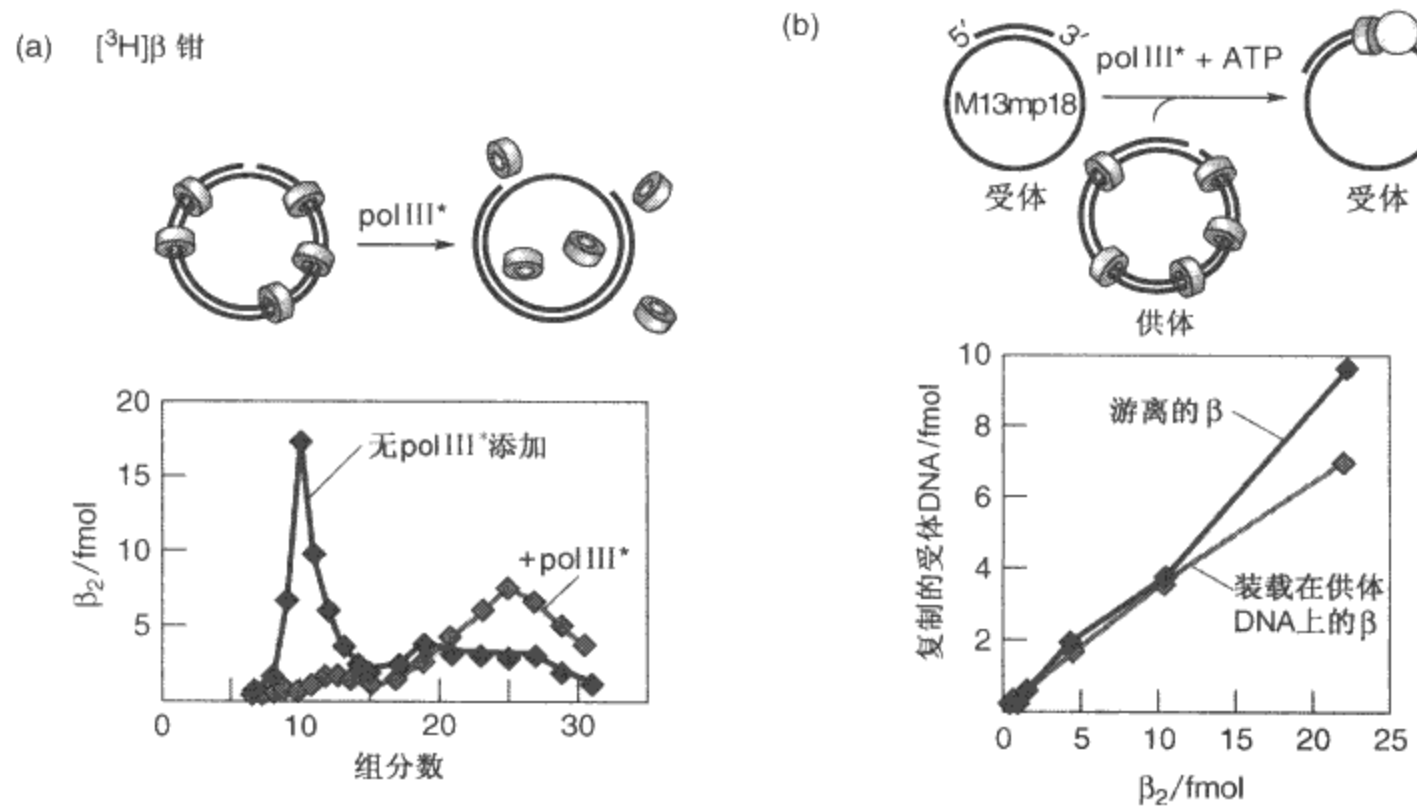


图 21.21　pol III* 具有钳卸载器的活性。（a）滑动钳的卸载。O'Donnell 及同事将用 γ 复合体将 β 钳（蓝色，上方）装载到有缺口的环形模板上，然后通过凝胶过滤去除 γ 复合体。加入 pol III* 后再次进行凝胶过滤。结果的曲线图（下方）显示，用 pol III* 处理后，β 钳（红色线）将从模板上释放出来，而未经 pol III* 处理，则 β 钳仍然与模板结合。（b）β 钳的再循环。经 pol III* 处理后，从 β 钳-模板复合体中释放出的 β 钳（红色线）与溶液中游离的 β 钳（蓝色线）一样，易与受体模板重新结合。（*Source*：Adapted from Stukenberg，P. T.，J. Turner and M. O'Donnell，An explanation for lagging strand replication：Polymerase hopping among DNA sliding clamps. *Cell* 78：883，1994.）

题，研究表明，核心聚合酶的α亚基、钳装载器的δ亚基均能与β亚基发生相互作用。研究者采用的分析方法是**蛋白质足迹法**（protein foot-printing），其原理与DNA足迹法相同，研究样品是标记的蛋白质而不是DNA，使用的是蛋白质切割试剂而不是DNA酶。在本实验中，O'Donnell及同事通过基因操作，在β亚基的C端引入一段可被蛋白质激酶识别的六氨基酸序列，改造后的产物命名为β^{PK}。在体外，研究者用蛋白激酶和标记ATP（γ-磷酸基团中的氧被^{35}S取代）对β^{PK}进行磷酸化修饰，使β^{PK}在C端被标记（与DNA足迹法中末端标记DNA类似）。研究者首先发现，钳装载器的δ亚基和核心聚合酶的α亚基均能保护β^{PK}，使之免于磷酸化修饰，这说明这两种蛋白质均能与β^{PK}发生直接相互作用。

蛋白质足迹分析强化了这些结论。O'Donnell及同事将标记的β^{PK}与不同的蛋白质混合，然后用链酶蛋白酶E和蛋白酶V8这两种蛋白水解酶切割蛋白质混合物，图21.22描述了实验结果。每图下

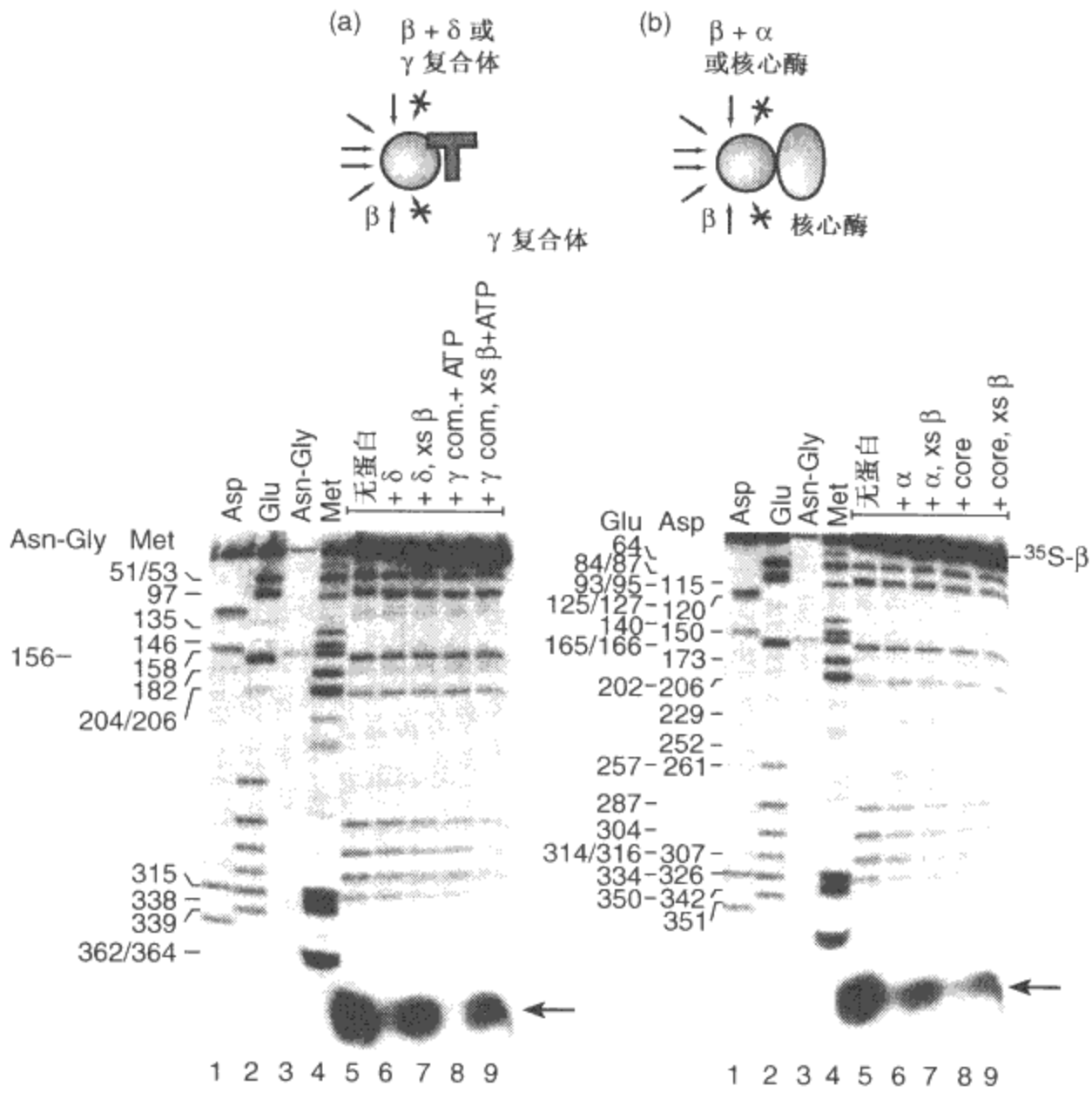


图21.22　β与γ复合体和核心聚合酶的蛋白足迹分析。O'Donnell及同事用蛋白激酶和［^{35}S］ATP对β^{PK}进行磷酸化修饰，使β^{PK}在C端被标记，将末端标记的β与δ亚基或整个γ复合体混合（a）或者与α或整个核心酶混合（b）。然后用链霉蛋白酶E和蛋白酶V8的混合物对蛋白质复合物进行中度切割，产生系列具末端标记的消化产物。最后，研究者进行电泳并放射自显影检测。每一图中的前4个泳道的消化产物作为分子标记，每一处理的特异性氨基酸标注在上方。泳道1：蛋白质经蛋白酶处理，该蛋白酶特异性消化天冬氨酸残基（Asp）。泳道5：无其他蛋白质存在时，对β^{PK}进行消化处理。泳道6～9：在每个泳道上所列蛋白存在时，切割β^{PK}的情况。结果δ和α亚基及γ和核心酶复合物均能保护相同的酶切位点。因而，凝胶底部箭头所示片段的量有所减少。上图展示了β钳与γ复合体（a）或核心酶（b）之间的结合，强调两者都结合在β钳中β单体近C端的同一区域，防止在此处切割（带"×"箭头所示）。（*Source*：Naktinis，V.，J. Turner and M. O'Donnell, A molecular switch in a replicating machine defined by an internal competition for protein rings. *Cell* 84（12 June 1996）f. 3ab bottoms，p. 138. Reprinted by permission of Elsevier Science.）

方中的前 4 个泳道是用 4 种已知切割位点的不同试剂切割标记的 β 亚基的结果。作为 Marker，每一图中的泳道 5 为在没有其他蛋白质存在时切割 β 亚基产生的末端标记肽，这些末端标记肽呈现梯度分布；图 21.22a 中的泳道 6 为在 δ 亚基存在情况下的结果，与泳道 5 一样，也表现为梯度分布，只是缺失了最小的片段（箭头所示），可能是丢失了或丰度大幅度降低。这表明 δ 亚基在 C 端附近与 β 亚基结合，阻止了蛋白酶在此位点进行切割。如果 δ 与 β 间的相互作用是特异的，则可以通过加入大量未标记的 β 亚基使之与 δ 亚基结合，从而阻断 δ 亚基与 β^{PK} 的结合，泳道 7 为这一实验的结果。图 21.22a 中的泳道 8 和 9 与泳道 6 和 7 类似，只是O'Donnell 及同事用 γ 复合体取代了纯化的 δ 亚基。同样，γ 复合体保护的位点也位于 β^{PK} 的 C 端附近，未标记的 β 亚基能阻止这种保护作用。

图 21.22b 与图 21.22a 十分相近，只是研究者在对标记的 β^{PK} 进行足迹分析时，添加的蛋白质是 α 亚基和整个核心酶，而不是 δ 亚基和整个 γ 复合体。研究者观察到了几乎一样的结果：α 亚基和整个核心酶以及 δ 亚基和整个 γ 复合体保护同一位点，这表明核心酶和钳装载器能够与 β 亚基的同一位点相结合，而 α 和 δ 亚基则分别介导了蛋白质之间的结合。在更为深入的实验研究中，因为 pol III* 包含核心酶和钳装载器，所以研究者用 pol III* 对 β^{PK} 做蛋白质足迹分析，预计可能会比分离的亚组分产生更多的足迹，然而事实并不是这样，这与 pol III* 通过核心酶或钳装载器而与 β 发生相互作用，但并不需要核心酶和钳装载器同时参与的假设相一致。

如果 β 钳既能与核心酶结合也能与钳装载器结合，但不能与两者同时结合，那么，β 钳将会优先与谁结合？O'Donnell 及同事用凝胶过滤的方法分析表明，当蛋白质在溶液中处于游离状态时，β 钳会优先与钳装载器结合，这是因为，在与核心聚合酶结合前，β 钳需要通过与钳装载器结合装载到 DNA 模板上。然而，一旦 β 钳被装载到有引物的 DNA 模板上时，情况将发生改变。当 β 钳被装载到 DNA 模板上后，β 需要与核心酶结合，起始 DNA 的合成。为验证这一假设，O'Donnell 及同事将 ^{35}S 标记的 β 钳装载到有引物的 M13 噬菌体 DNA 上，然后分别添加标记的钳装载器（γ 复合体）和未被标记的核心酶（图 21.23a）或 ^{3}H 标记

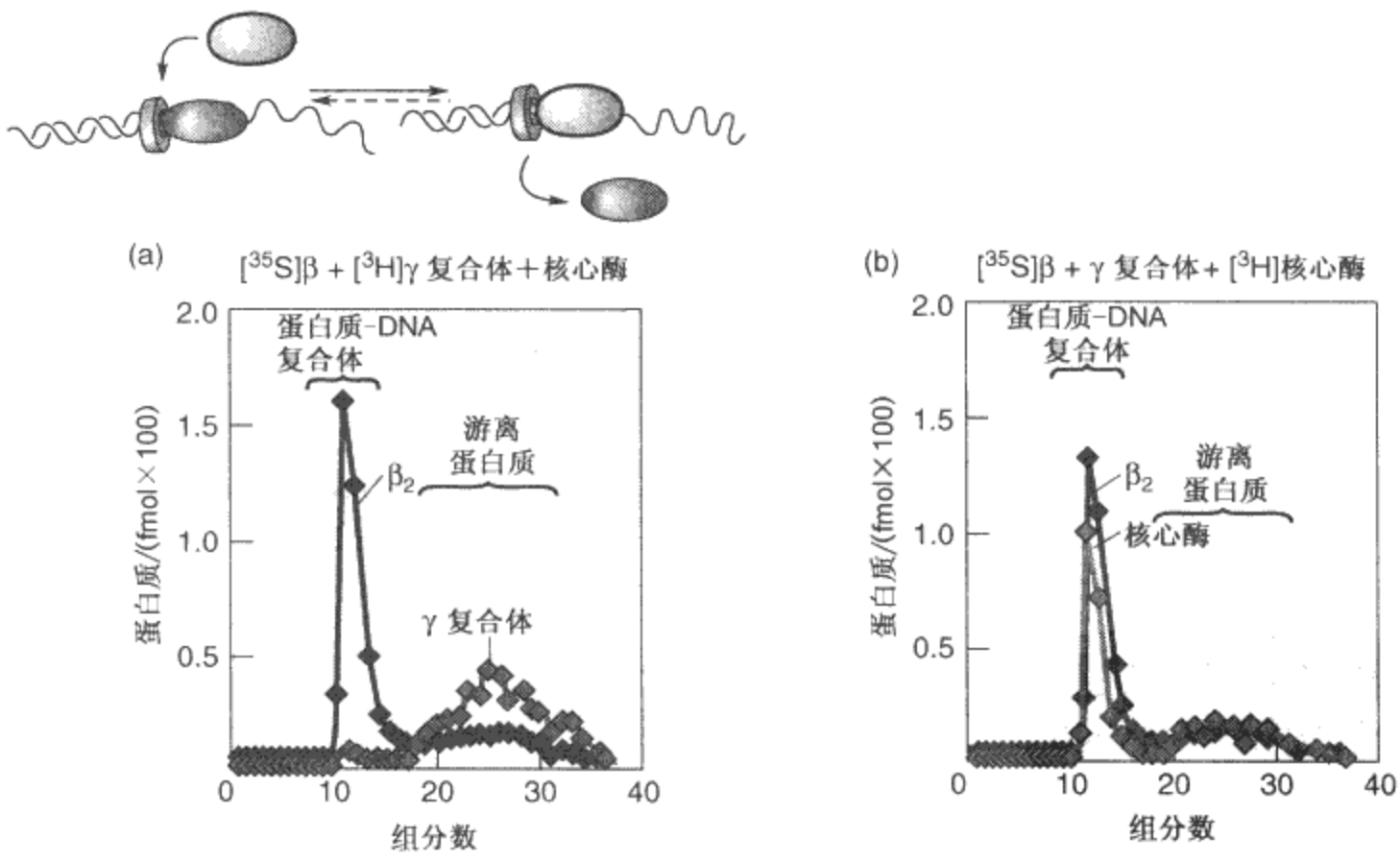


图 21.23　在结合有引物的模板存在时，β 钳优先结合核心酶而不是 γ 复合体。O'Donnell 及同事将 β 钳（^{35}S 标记）装载到具引物的 M13 噬菌体 DNA 模板上，如图左上方所示。然后加入 ^{3}H 标记的 γ 复合体（红色）及未标记的核心酶（绿色），将混合物进行凝胶过滤处理，以检测标记的 γ 复合体是否仍与 β 钳-DNA 复合体结合在一起。图（a）所示，标记的 γ 复合体（红色）独立于 β 钳-DNA 复合体而析出。此外，研究者加入 ^{3}H 标记的核心酶和未标记的 γ 复合体，重复进行凝胶过滤处理。如图（b）所示，标记的核心酶（绿色）与 β 钳-DNA 复合体（蓝色）共析出，说明当 β 钳与具引物的 DNA 模板结合时，核心酶将与其结合。（*Source*：Adapted from Naktinis，V.，J. Turner and M. O'Donnell，A molecular switch in a replication machine defined by an internal competition for protein rings. *Cell* 84：140，1996.）

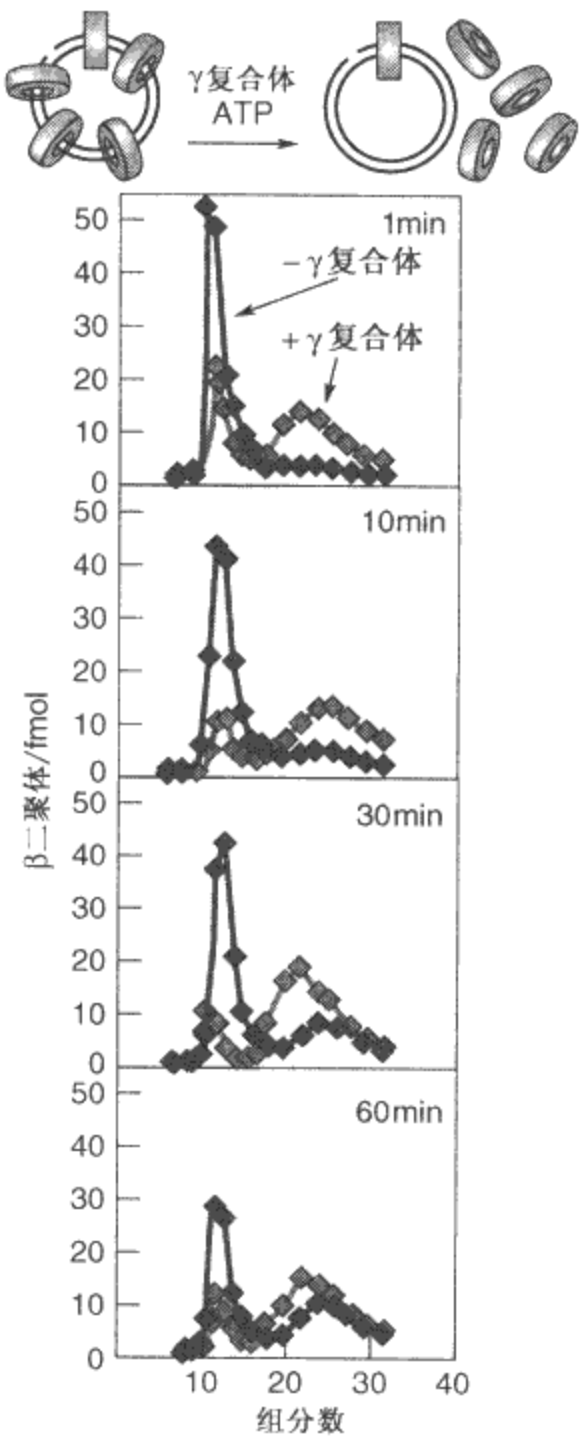


图 21.24　γ 复合体的钳卸载器活性。O'Donnell 及同事将 β 钳装载到有缺口的环状 DNA 模板上，如图上方所示。然后在 γ 复合体和 ATP 存在（红色）或不存在（蓝色）两种情况下，按图中所示的时间孵育复合物。最后将混合物进行凝胶过滤处理，以检测还有多少 β 钳仍与 DNA 结合，有多少已经解离。图上方的卡通形象地阐释了如下结果：γ 复合体和 ATP 加速了 β 钳从缺口 DNA 上的卸载。（*Source*：Adapted from Naktinis，V.，J. Turner and M. O'Donnell，A molecular switch in a replication machine defined by an internal competition for protein rings. *Cell* 84：141，1996.）

的核心酶和未被标记的 γ 复合体（图 21.23b），将混合物进行凝胶过滤处理，从游离的蛋白质中分离 DNA-蛋白质复合体。在这些条件下，装载在 DNA 模板上的 β 钳显然会优先与核心聚合酶结合，在 β 钳-DNA 复合体上几乎没有 γ 复合体的结合。

一旦完成一个冈崎片段的合成，全酶必须与 β 钳解离，然后移向新的 β 钳。而原来的 β 钳必须从模板上释放下来，以便参与另一个冈崎片段的合成。我们虽然已知 pol III* 具有钳卸载活性，但并不了解是 pol III* 的哪一部分具有卸载活性。O'Donnell 及同事通过凝胶过滤实验分析表明，γ 复合体具有钳卸载器活性。图 21.24 描述了该项实验的操作过程与结果。研究者将 β 钳装载到有缺口的 DNA 模板上，去除其他所有蛋白质，然后在 γ 复合体存在或不存在两种情况下，孵育 DNA-蛋白质复合体。研究发现，在 γ 复合体和 ATP 存在时，β 钳会很快地从缺口 DNA 模板上被卸载下来，而 γ 复合体和 ATP 不存在时，卸载速度较慢。

γ 复合体既是钳装载器，又是钳卸载器。但是，由什么机制能够决定 γ 复合体何时装载或卸载滑动钳呢？如图 21.25 所示，DNA 的状态似乎是一个开

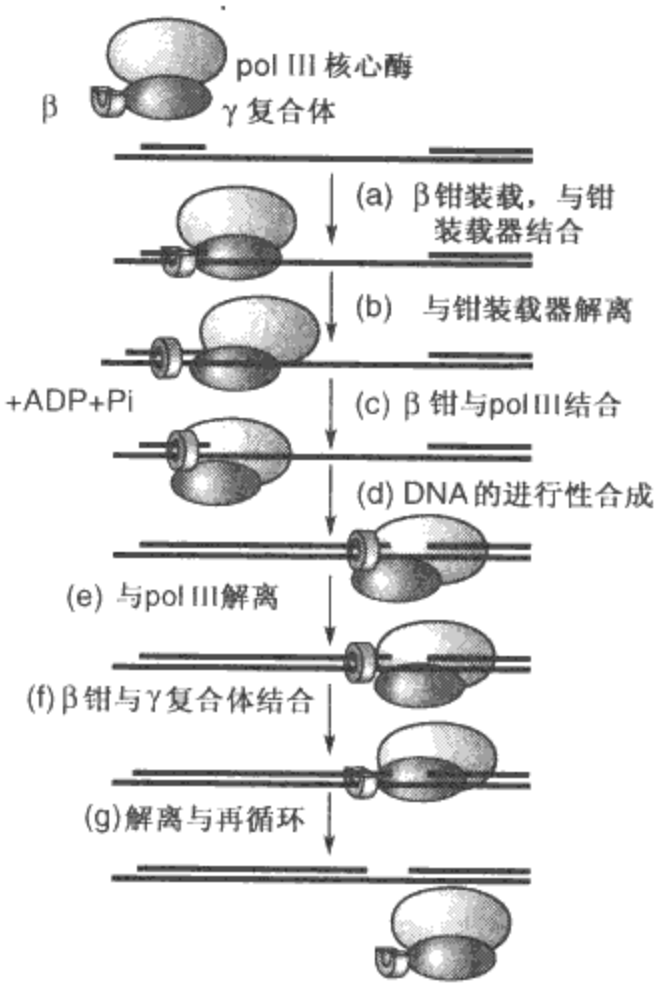


图 21.25　后随链复制概述。从 β 钳与 pol III* 的 γ 复合体（红色）开始。（a）γ 复合体将 β 钳（蓝色）装载到结合有引物的模板上。（b）γ 复合体或钳卸载器与 β 钳解聚。（c）核心酶（绿色）与 β 钳结合。（d）核心酶与 β 钳协同进行性地合成冈崎片段，在两个冈崎片段之间会留有缺口。（e）聚合酶核心与 β 钳解离。（f）γ 复合体与 β 钳重新结合。（g）γ 复合体作为钳卸载器，使 β 钳脱离模板。此时，游离的 β 钳可以进入新一轮循环，再循环到模板的另一个引物处。（*Source*：Adapted from Herendeen，D. R. and T. J. Kelly，DNA polymerase III：Running rings around the fork. *Cell* 84：7，1996.）

关。当β钳以游离形式存在于溶液中且有结合有引物的模板可供利用时，β钳会优先与γ复合体结合，此时，γ复合体作为钳装载器将β钳装载到DNA模板上。一旦与DNA结合，β钳将优先与核心聚合酶结合，使之在合成冈崎片段时具有很高的进行性。当片段合成、只留下一个缺口后，核心聚合酶便失去与β钳的亲和力。于是，β钳重新与γ复合体结合，γ复合体作为钳卸载器将β钳从DNA模板上释放下来，使之再循环到下一个引物处，起始新一轮的复制。

小结 pol III全酶的2个核心聚合酶通过2个τ亚基与γ复合体连接在一起。其中一个核心聚合酶负责前导链的连续合成，另一个核心聚合酶负责后随链的不连续合成。γ复合体作为钳装载器将β钳装载到结合有引物的DNA模板上。一旦完成装载，β钳失去与γ复合体的亲和力，转而与核心聚合酶结合，帮助核心聚合酶高度进行性地合成冈崎片段。一旦完成冈崎片段的合成，β钳失去与核心聚合酶的亲和力，进而再与γ复合体结合，γ复合体作为钳卸载器将β钳从DNA模板上释放下来，使之再循环到下一个引物处，起始新一轮的复制。实现此过程的重复。

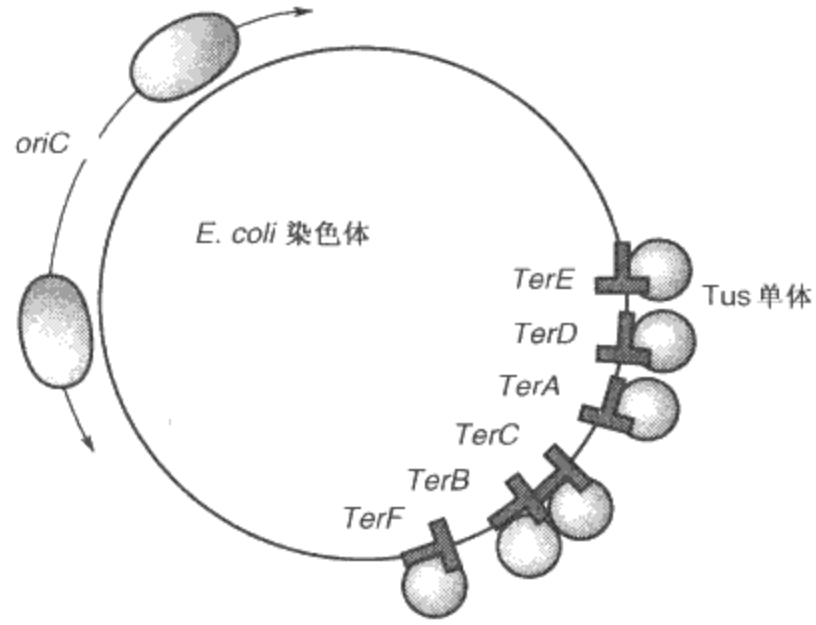


图21.26 *E. coli*基因组的终止区。两个携带复制体（绿色）的复制叉从*oriC*开始呈相反方向向环状*E. coli*染色体的终止区移动。3个终止子位点调控一个复制叉：*TerE*、*TerD*和*TerA*终止逆时针方向移动的复制叉；*TerF*、*TerB*和*TerC*终止顺时针方向移动的复制叉。Tus蛋白结合在终止子位点上，帮助终止子位点阻止移动的复制叉。（*Source*：Adapted from Baker, T. A., Replication arrest. *Cell* 80: 521, 1995.）

21.3 复制的终止

对于λ噬菌体及其他一些噬菌体而言，通过复制会产生长的线性连环体，复制的终止相对简单。持续复制的连环体被按基因组大小切割，然后被包装到噬菌体的头部。但对细菌和真核生物而言，其DNA复制有明确的起点和终点，复制终止的机制更为复杂，也更值得我们去探究。细菌DNA在复制过程中，2个复制叉在终止区相互靠近，终止区内包含有能够与特异蛋白结合的22bp的终止子位点。*TerA-TerF*是*E. coli*的终止子（*Ter*）位点，其排列次序如图21.26所示。*Ter*位点结合Tus蛋白（终止相关蛋白）。复制叉进入到终止区后，在彻底完成复制过程前会发生暂停，导致2个子代双链DNA分子缠绕在一起。缠绕在一起子代双链DNA分子必须在细胞分裂前解除缠绕，否则就不能被分配到2个子代细胞中。如果不能解除缠绕，子代双链DNA分子就会停留在细胞的中间位置，细胞分裂失败并可能会导致细胞的死亡。我们不禁会问：子代双链DNA分子如何解除缠绕？对真核生物而言，我们希望知道细胞是如何补平因引物清除而在线性染色体5′端留下的缺口。下面我们将对这些问题逐一进行分析。

解连接：解开缠绕的子代DNA分子

图21.27描述了细菌在DNA复制即将结束时所面临的问题。由于细菌DNA分子为环形，2个子代双螺旋分子如2个相扣的环一样缠绕在一起，形成环链体（catenane），为使相扣的DNA分子分配到子代细胞中，必须解除连接。如果解连接发生在修复合成之前，那么，单个切口就足以使DNA解除连接，拓扑异构酶I可以解除连接。然而，如果修复合成先行发生，则需要拓扑异构酶II完成解连接，该酶能够使DNA双螺旋穿过双链断口。鼠伤寒沙门氏菌（*Salmonella typhimurium*）和*E. coli*细胞有4种拓扑异构酶：拓扑异构酶I～IV（topo I～IV）。topo I和topo III属于I型拓扑异构酶，topo II和topo IV属于II型拓扑异构酶。问题是：哪种拓扑异构酶参与子代DNA分子的拆分？

由于DNA旋转酶（topo II）在DNA复制过程中发挥解旋的功能，因此许多分子生物学家猜测DNA旋转酶也能使子代双链分子解除连接。但是，Nicholas Cozzarelli及同事证实拓扑异构酶IV才是真正的解连酶。在体外许可和非许可温度条件下，研究者测试了鼠伤寒沙门氏菌（*E. coli*的近缘菌）不同温度敏感突变菌株解开质粒pBR322二聚体的

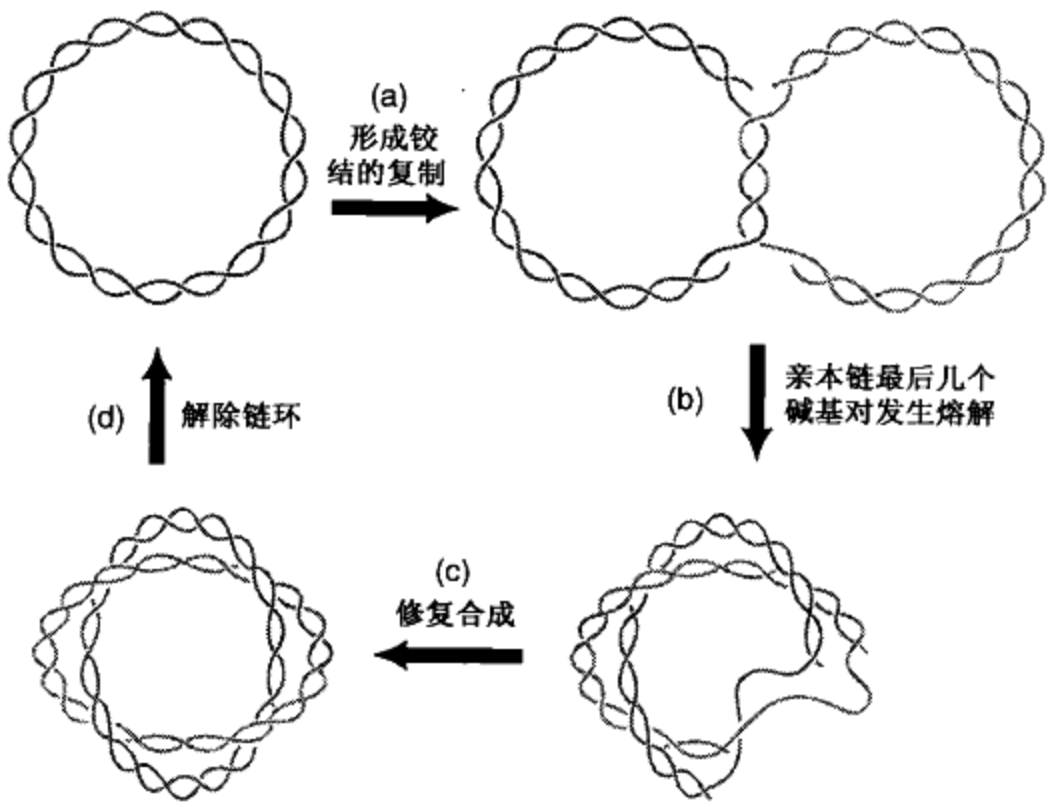


图 21.27　解除链环的两阶段模型。(a) 当环状细菌 DNA 进行 θ 模型复制时，利用了 DNA 旋转酶的旋转活性，当复制停止后，亲本链上还保留一些缠绕的双螺旋。(b) 留存的亲本双螺旋熔解，子代 DNA 分子中伸展的 DNA 单链相互缠绕。(c) 修复合成填平单链缺口，但两条子双链仍然缠绕成链环体。链环体中的扭结数或一个双螺旋分子与另一个双螺旋分子相交的次数，与经过步骤 (a) 后未被复制的亲本双螺旋的交叉次数相同，本图为 4 个。(d) 2 个子代双链 DNA 分子分离。(*Source*: Adapted from Adams, D. E., E. M. Shekhtman, E. I. Zechiedrich, M. B. Schmid and N. R. Cozzarelli, The role of topoisomerase IV in partitioning bacterial replicons and the structure of catenated intermediates in DNA replication. *Cell* 71: 278, 1992.)

能力。研究发现，*gyrA*、*parC*、*parE*、*parF* 等基因发生了突变。*gyrA* 编码 topo II 的 A 亚基；*parC* 和 *parE* 编码 topo IV 的 2 个亚基；*parF* 与 *parC* 和 *parE* 连锁。研究者还测试了 DNA 旋转酶的潜在抑制剂诺氟沙星 (norfloxacin) 的影响效应，因为诺氟沙星是 DNA 旋转酶的潜在抑制剂，因此也是 DNA 复制的抑制剂。

为分析环的解离，研究者在诺氟沙星存在或不存在两种情况下，分别于 30℃或 44℃条件下培养细胞，然后从细胞中提取 DNA。用 DNase I 处理使 DNA 分子产生切口，以松弛超螺旋，电泳后进行 Southern 斑点杂交。最后，pBR322 探针与 Southern 斑点进行杂交。在 Southern 斑点杂交的放射自显影图像上，环连体呈现系列条带，分布于切口二聚体和切口单体标记之间，每一条带对应的环连体具有不同的节点数或交叉数（如图 21.27 左下方的结构有 4 个节点，在节点处，2 个质粒彼此交叉）。值得注意的是，节点数越多，环连体的电泳迁移率越大。图 21.28 中放射自显影图像显示，*parC* 和 *parE* 基因发生损伤的突变体在无诺氟沙星存在并且非许可温度 (44℃) 条件下不能将质粒解连环，这表明 topo IV 对解连接是十分重要的。诺氟沙星通过阻断 DNA 旋转酶活性而使 DNA 复制发生暂停，并且很可能允许少量的 topo IV 或其他拓扑异构酶进行后续的解连接。相反，DNA 旋转酶发生突变的菌株在非许可温度条件下，不管诺氟沙星是否存在，都不会积聚环连体，这暗示 DNA 旋转酶没有参与解连接。在检测 *E. coli* 温度敏感型 *parC* 和 *parE* 突变体时，Cozzarelli 及同事观察到了类似的现象，说明在 *E. coli* 中，topo IV 也参与了解连接。

图 21.27 左下方显示的环连体有 4 个节点，因为右上方的 2 个子代双螺旋 DNA 分子在复制即将结束时还有 4 个交叉点。图 21.27 描述的两阶段机制提出了一些假设。第一，在我们称之为螺绕环 (torus)（螺绕环为环形结构）的环连体中，2 个子代质粒彼此螺旋缠绕；第二，由于 DNA 是右手螺旋，所以，2 个子代质粒彼此缠绕形成的螺旋也一定是右手螺旋；第三，我们观察到的节点数目应为偶数；最后，由于 2 个子代质粒具有相同的方向，所以环连体应该是平行结构。只有两阶段模型才能形成右手平行的螺绕环连体。

为验证温度升高后突变细胞内剩余环连体的特征，Cozzarelli 及同事用 RecA 蛋白将环连体包装成易于观察的形式，于电子显微镜下观察。研究者观

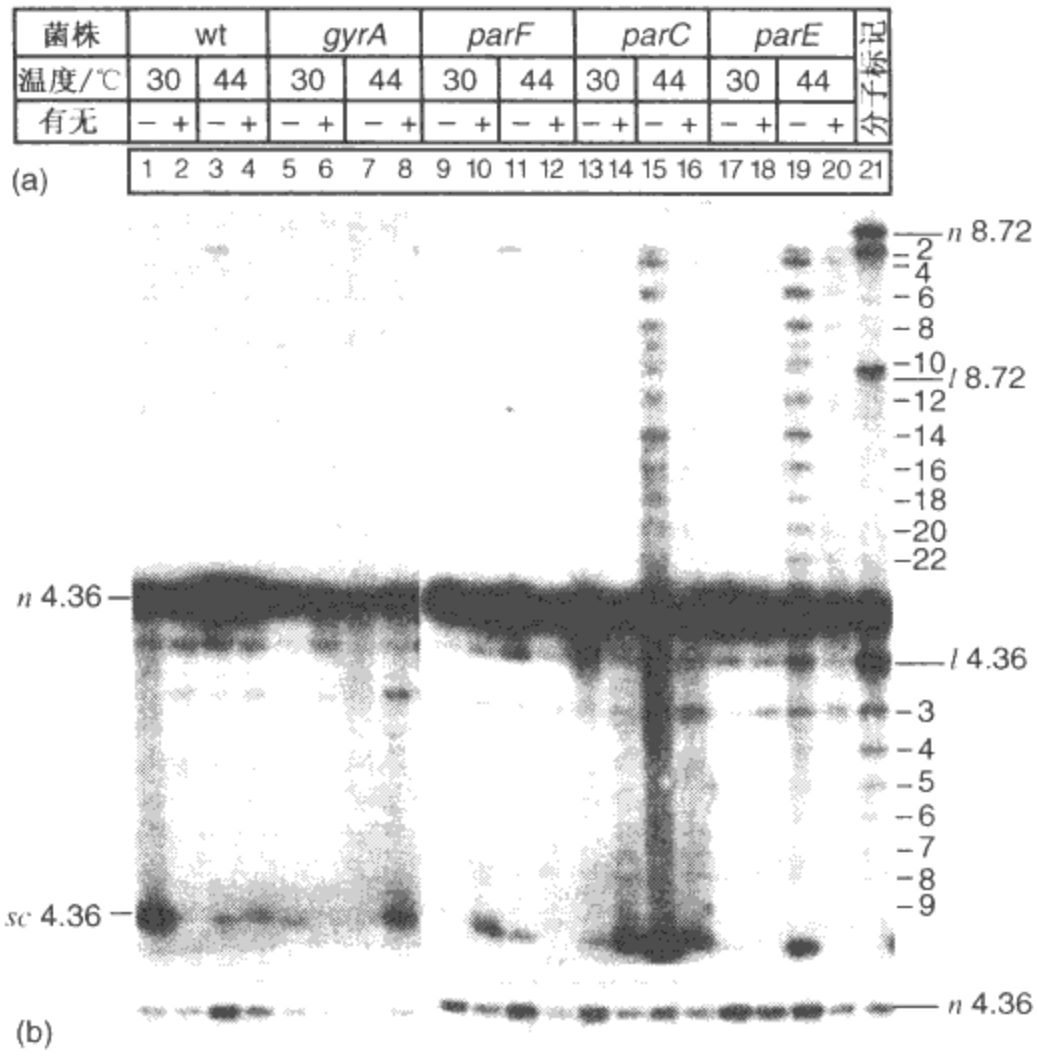


图 21.28　鼠伤寒沙门氏菌突变体中环连质粒的积累。（a）Southern 斑点杂交的放射自显影图像。Cozzarelli 及同事分别在许可温度（30℃）或非许可温度（44℃）条件下，在存在或不存在诺氟沙星的培养基上培养野生型及 4 种温度敏感型沙门氏菌突变菌株，如图上方所示。所有菌株都携带 pBR322 质粒。从培养的细菌细胞中分离质粒 DNA，同过引入切口松弛超螺旋，进行琼脂糖凝胶电泳分离节点数不同的环连体。进行 Southern 斑点杂交后，用标记的 pBR322 DNA 为探针检测 DNA 斑点。标为 *n* 和 *l* 的 Marker 为有切口的线性质粒 DNA，其大小（kb 为单位）分别标注图的右侧和左侧。sc 表示超螺旋质粒的位置。*parC* 和 *parE* 突变株在非许可温度下不能将连接的子代质粒解开，但是其他菌株包括温度敏感型 DNA 解旋酶 *gyrA* 突变菌株能够在非许可温度下解开连接。（b）对切口质粒 pBR322（n 4.36）的斑点进行曝光不足处理，以显示每个泳道的加样量。（*Source*：（a-b）Adams，D. E.，E. M. Shekhtman，E. L. Zechiedrich，M. B. Schmid and N. R. Cozzarelli，The role of topoisomerase IV in partitioning bacterial replicons and the structure of catenated intermediates in DNA replication. *Cell* 71（16 Oct 1992）f. 2，p. 279. Reprinted by permission of Elsevier Science.）

察到多种具有偶数节点的螺绕环连体结构（图 21.28 也表明，在环连体中只有偶数个节点），而且，仔细查验显微图像发现，正如两阶段模型所预测的那样，螺绕环连体为右手螺旋。更为复杂的分析表明 2 个子代质粒确实是平行的。这样，两阶段模型所需的所有条件就完全符合，相关数据也有力地支持了该模型，即 topo IV 对质粒具有解连接作用。但是 topo IV 在细菌染色体的解连接中是否具有相同的功能呢？答案是肯定的，因为 topo IV 突变是致死的，死亡细胞内存在未完全分离的染色体。

真核生物染色体不是环形，但存在多个复制子，因此，来自相邻复制子的复制叉彼此靠近时，与细菌染色体的 2 个复制叉在终止位点彼此靠近的情形十分相近。显然，这会阻止 DNA 复制的最终完成，所以，真核生物的染色体也会形成必需被解连接的环连体。真核生物的 topo II 与细菌 DNA 旋转酶相比更类似于 topo IV，是解连酶的重要候选者。

小结　在复制过程的末期，环状细菌染色体形成了需通过两步过程才能解离的环连体。首先，连接两条链未被复制的双螺旋熔解，然后修复合成填补空隙。形成了具有偶数节点的右手平行螺绕环连体，由拓扑异构酶 IV 从节点处将环连体解开。线性真核生物染色体在 DNA 复制过程中也需要解连接。

真核生物 DNA 复制的终止

真核生物在 DNA 复制结束时面临着原核生物

不会遇到的困难：填补 RNA 引物去除后留下的空隙。对细菌等的环形 DNA 而言，填补所有的空隙是没有任何问题的，因为在空隙的上游总有其他 DNA 片段的 3′端起到引物的作用（图 21.29a）。但是，由于真核生物染色体为线性结构，一旦每条链上的第一个引物被去除（图 21.32b），由于 DNA 分子不能按 3′→5′方向延伸，也不像环形 DNA 分子那样拥有上游 3′端，所以，留下的空隙是无法填补的。如果情况确实如此，这样，DNA 每复制一次染色体就会缩短一段 RNA 引物的长度。那么细胞将如何解决这个问题？

端粒维持　Elizabeth Blackburn 及同事找到了问题的答案，总结如图 21.29c。真核生物染色体末端的**端粒**（telomere）由短的富含 GC 的重复序列组成。端粒中富含 G 的链在**端粒酶**（telomerase）的作用下以非半保留的方式被添加在 DNA 链的 3′端。端粒中的重复序列具有种属特异性，在四膜虫（*Tetrahymena*）中，该序列为 TTGGGG /AACCCC；在包括人在内的脊椎动物中，端粒序列为 TTAGGG/AATCCC。Blackburn 指出，这种特异性取决于端粒酶自身，因为端粒的合成需要以端粒酶的小 RNA 分子为模板。这样就解决了末端短缩的问题：端粒酶在染色体的 3′端添加具有自身序列特征的许多重复序列，然后在端粒内引发富含 C 链的合成。即使引物被去除且无 DNA 序列填补也无关紧要，因为只是端粒序列发生了丢失，而且丢失的序列可通过端粒酶及新一轮的端粒合成所恢复。

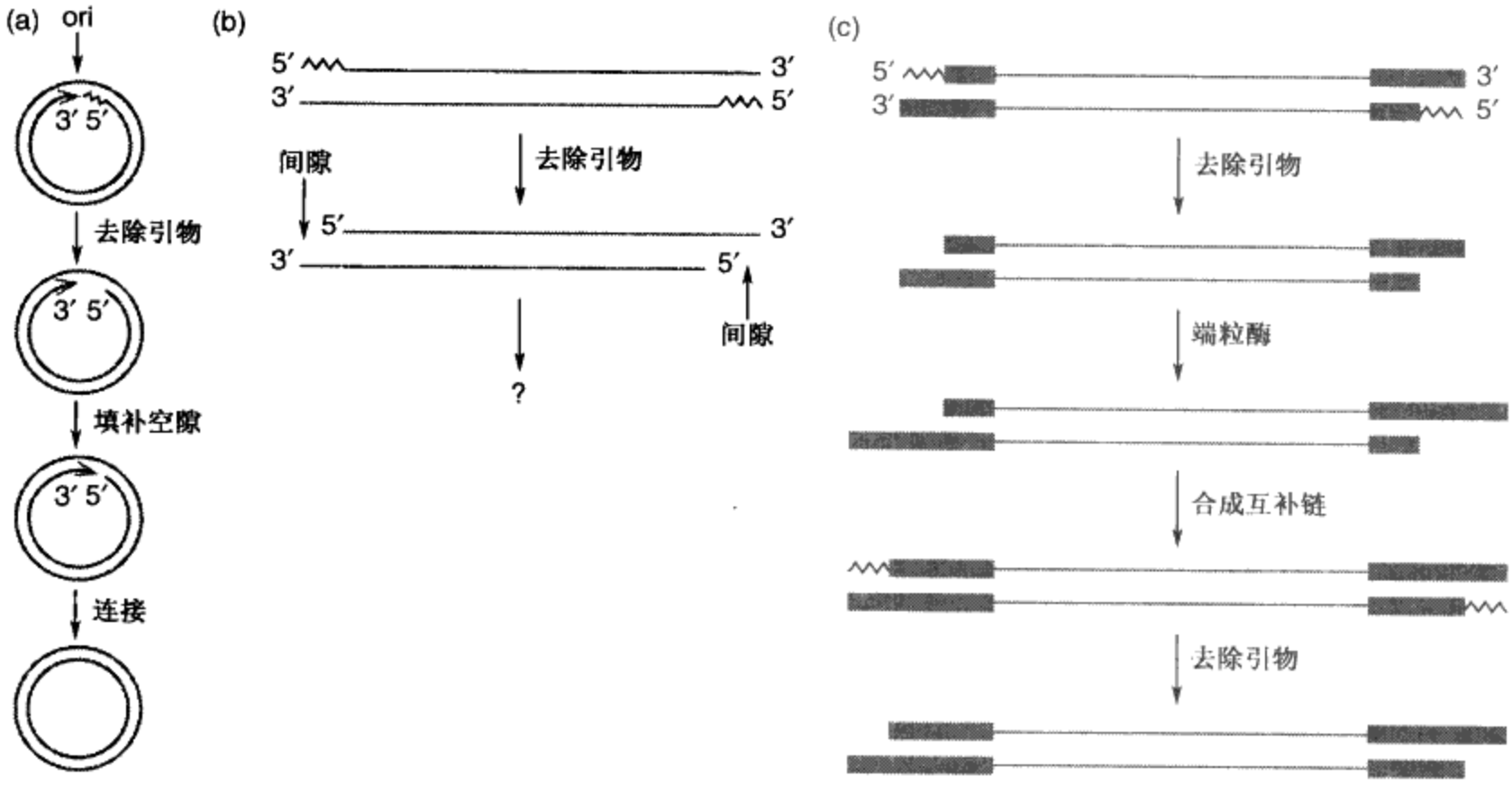


图 21.29　引物去除后残留缺口的处理。（a）在细菌中，环状 DNA 链的 3′端可作为引物而引发 DNA 的合成，填补第一条引物（红色）切除后残留的缺口。为简便起见，图中只显示了一条复制的链。（b）假设模型。如果线性 DNA 链 5′端只是简单地将引物去除，而没有端粒酶发挥作用，那么，染色体末端的缺口在每一次 DNA 复制后都会增长。（c）端粒酶解决末端端缩的方式。第一步，子链 5′端的引物（红色）被去除，留下缺口。第二步，端粒酶在另一条子链的 3′端加上额外的端粒 DNA（绿色框）。第三步，以新加入的端粒 DNA 作为模板进行 DNA 合成。第四步，第三步使用的引物被去除。留下缺口，但是端粒酶活性确保不会发生 DNA 缺失。图中所示的端粒与引物的长度并不成比例。人的端粒可长达几千个核苷酸。（*Source*：（c）Adapted from Greider, C. W. and E. H. Blackburn, Identification of a specific telomere terminal transferase activity in tetramere extracts. *Cell* 43 (Dec Pt1 1985) f. 1A, p. 406.）

Blackburn 很明智地选择纤毛类原生动物——四膜虫（*Tetrahymena*）来研究端粒酶活性。四膜虫有两类细胞核：①小核，含有全部基因组的 5 对染色体，负责基因由亲代向子代的传递；②大核，5 对染色体断裂成 200 多条短片段，负责基因的表达。这些微染色体的末端均携带有端粒，所以四膜虫细胞的端粒要多于人类细胞，这些端粒都在端粒酶的作用下添加到微染色体的末端，特别是在大核发育阶段，新形成的微染色体必须在端粒酶的作用添加端粒，因此，从四膜虫细胞中分离端粒酶比较容易。

1985 年，Carpl Greider 和 Blackburn 通过对处于大核发育阶段的同步化四膜虫细胞提取物的研究.

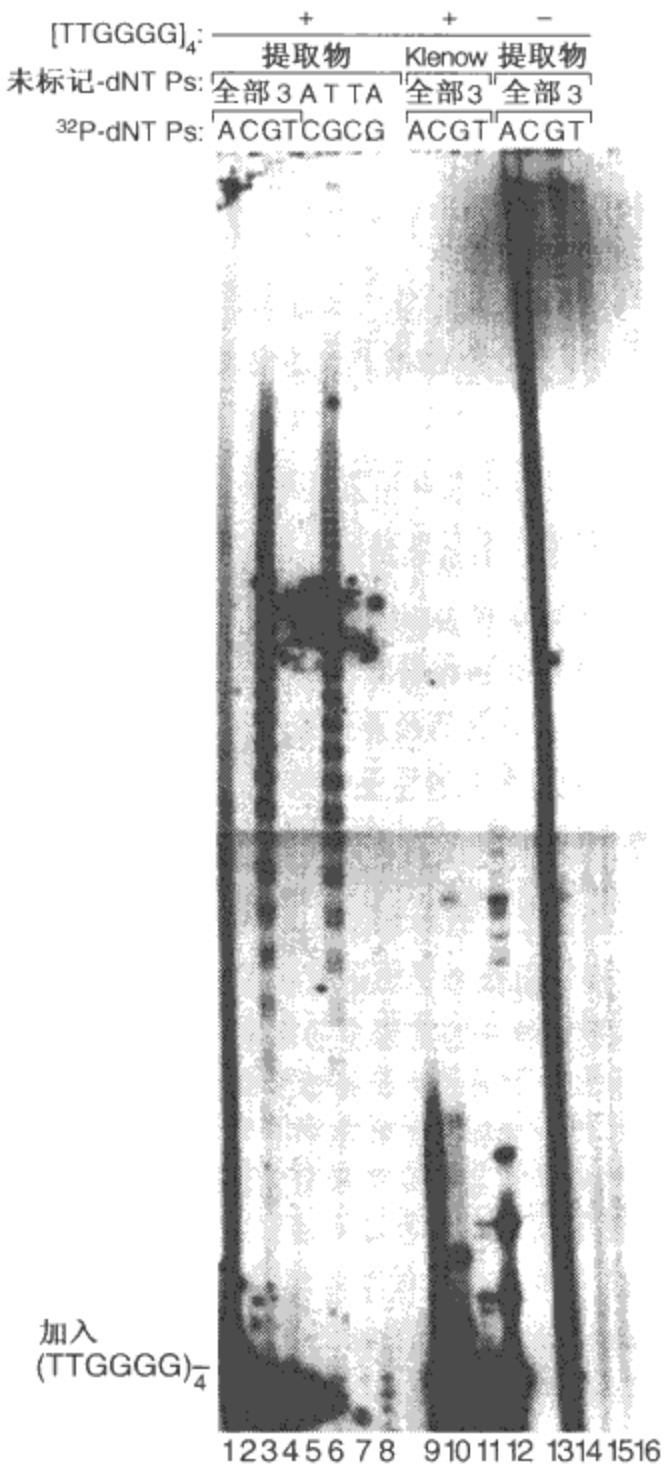


图 21.30 端粒酶活性的鉴定。Carpl Greider 和 Blackburn 使四膜虫细胞同步交配，并使子代生长到大核发育阶段。研究者制备无细胞提取物，在加入标记和未标记核苷酸的情况下（图上方所示），与人工合成的有 4 个 TTGGGG 端粒重复序列的寡聚体共同温育 90min，产物经电泳后进行放射自显影检测。泳道 9～12 为 *E. coil* DNA 聚合酶 I 的 Klenow 片段替换了四膜虫细胞提取物。泳道 13～16 为细胞提取物，但没有引物。端粒酶显然只有在 dGTP 和 dTTP 都存在的情况下才有活性。（*Source*：Greider，C. W. and E. H. Blackburn，Identification of a specific telomere terminal transferase activity in tetramere extracts. *Cell* 43（Dec Pt1 1985）f. 1A，p. 406. Reprinted by permission of Elsevier Science.）

成功鉴定了端粒酶的活性。在体外，研究者利用包含 4 个重复的 TTGGGG 端粒序列的合成引物分析端粒酶的活性，反应体系中添加放射性核苷酸以标记延伸的端粒样 DNA 产物。图 21.30 显示了实验的结果，泳道 1～4 为添加不同标记核苷酸（分别为 dATP、dCTP、dGTP 和 dTTP）、同时添加其他 3 种未标记核苷酸的实验结果，泳道 1 为添加标记 dATP 的结果，只出现一弥散片状条带；泳道 2 和 4 显示人工合成的端粒，没有获得延伸；泳道 3 为添加标记 dGTP 的实验结果，显示人工合成的端粒被周期性延伸，每一簇条带代表新添加了一至多个 TTGGGG 序列（由于完成程度的不同），所以呈现条带簇，而不是单条带。当然，同使用 dGTP 一样，使用 dTTP 也应该会使端粒延伸。进一步研究表明，本实验中使用的 dTTP 浓度太低，若在高浓度条件下，dTTP 能够掺入到端粒中。泳道 5～8 为添加一种标记核苷酸且只添加一种非标记核苷酸的实验结果。正如所料，该实验证实只有在未标记 dTTP 存在条件下，dGTP 才能掺入到端粒中，因为这条端粒链仅由 G 和 T 组成。泳道 9～12 为对照，表明常规的 DNA 聚合酶、Klenow 片段不能延伸端粒；泳道 13～16 表明端粒酶活性依赖于端粒样的引物。

在没有互补 DNA 链可阅读的情况下，端粒酶如何在端粒末端添加正确的碱基序列？端粒酶在合成端粒时，是以自身的 RNA 组分为模板的（注意：RNA 作为模板而不是引物）。在 1987 年，Greider 和 Blackburn 发现，端粒酶是由端粒酶 RNA 和蛋白质亚基组成的核糖核蛋白。1989 年，他们克隆并测序了编码四膜虫端粒酶 159nt RNA 亚基的基因，该亚基包含有 CAACCCCAA 序列。理论上讲，该序列可以作为模板，在四膜虫端粒末端添加额外的重复序列 TTGGGG（图 21.31）。

Blackburn 及同事用遗传学方法证明了端粒酶 RNA 确实是端粒合成的模板。研究表明，端粒酶 RNA 发生突变，会引起端粒序列发生相应的改变。他们改变了四膜虫端粒酶 RNA 编码基因中 5′-CAACCCCAA-3′的序列：

wt：5′-CAACCCCAA-3′
1：5′-CAA<u>C</u>CCCCAA-3′
2：5′-CAACC<u>T</u>CAA-3′
3：5′-C<u>G</u>ACCCCAA-3′

在这 3 种突变体（1、2 和 3）中，带下画线的碱基表示该碱基发生了改变（或插入，如突变体 1）。研究者通过质粒将野生型或突变型基因导入四膜虫细胞，以保证基因能够超表达。虽然每种情况下都保留有内源野生型基因，但由于外源基因的超表达，内源野生型基因的效应将被掩盖。从每一类转化细胞中提取端粒 DNA，Southern 斑点杂交的结果显示，预期与突变体 1（TTGGGGG）和 3（GGGGTC）的

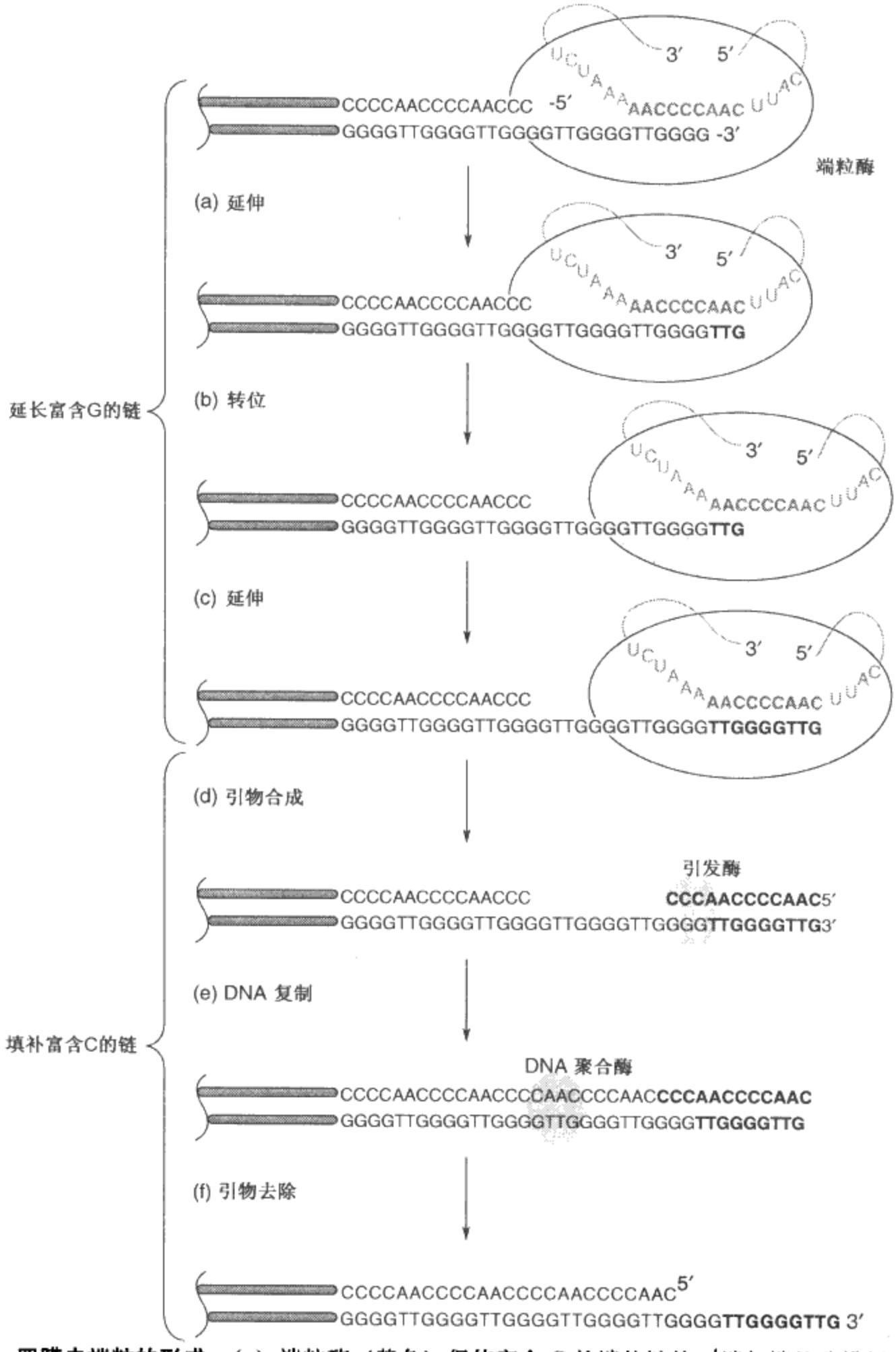


图 21.31　四膜虫端粒的形成。（a）端粒酶（黄色）促使富含 G 的端粒链的 3′端与端粒酶模板 RNA（红色）杂交。端粒酶利用自身 RNA 的 3 个碱基（AAC）为模板，在端粒的 3′端添加 3 个碱基（TTG，黑体）。（b）端粒酶移位到端粒新的 3′端，使其模板 RNA 左侧的 AAC 序列与端粒中新加入的 TTG 配对。（c）端粒酶利用模板 RNA 在端粒 3′端再加上 6 个核苷酸（GGGTTG，黑体）。从（a）至（c）的步骤能够被无限重复，从而使富含 G 的端粒链延长。（d）当富含 G 的链延伸至足够长时（可能比图示的要长），引发酶（橙色）合成能与富含 G 的端粒链的 3′端互补的 RNA 引物（黑体）。（e）DNA 聚合酶（绿色）利用新合成的引物引发 DNA 合成，填补存留在富含 C 的端粒链上的缺口。（f）引物被切除，在富含 G 的端粒链的 3′端形成长为 12～16nt 的突出端。

端粒序列杂交的探针确实与来自转化突变基因的细胞的端粒 DNA 发生了杂交。此外，这对突变体 2 不起作用，因为不存在能够与探针杂交的端粒 DNA，而该探针能够与 GAGGTT 序列杂交。

这些研究结果表明，突变的端粒酶 RNA1 和 RNA3 可以作为端粒延伸的模板，而突变的端粒酶 RNA2 则不能作为端粒延伸的模板。为验证这一推断，Blackburn 及同事转化端粒酶 RNA3 突变基因，对来自此类细胞的端粒片段进行测序，其序列如下：

5′-CTTTTACTCAATGTCAAAGAAATTAT-TAAATT$(GGGGTT)_{30}(GGGGTC)_2GGGGTT(GGGGTC)_8GGGGTTGGGGTC(GGGGTT)_N$-3′

带下画线的碱基一定是由突变的端粒酶 RNA 所编码。这个非均质序列与该物种十分一致的正常端粒序列存在很大差异。最初的 30 个重复序列似乎是在转化前已由野生型端粒酶 RNA 编码合成，其后是 11 个突变重复序列，其间包含有 2 个野生型重复序列，最后的序列均为野生型序列。末端的野生型序列可能是与野生型端粒重组的结果，或是在细胞丢失突变的端粒酶 RNA 后端粒才开始合成的结果。尽管如此，如果端粒序列是以突变端粒酶 RNA 为模板合成的，那么就应该存在大量重复的正确序列，事实也是如此，正如我们预料的那样。所以，端粒酶 RNA 是端粒合成的模板，正如图 21.31 所示。

端粒酶以 RNA 为模板合成 DNA 链的事实暗示端粒酶作为反转录酶而起作用。所以，Blackburn 及其他研究者开始着手纯化端粒酶，以研究分析端粒酶确实是以反转录酶的形式发挥功能的。在研究纯化端粒酶 10 年后的 1996 年，Joacim Lingner 和 Thomas Cech 最终成功地从另一种纤毛类原生动物游仆虫（*Euplotes*）中获得了纯化的端粒酶。该酶包含有 p43 和 p123 蛋白以及作为模板延伸端粒的 RNA 亚基。p123 蛋白具有反转录酶的特征序列，说明该蛋白为端粒酶提供了催化活性，所以又称之为**端粒酶反转录酶**（telomerase reverse transcriptase，**TERT**）。由于端粒酶反转录酶发现之时，恰是人类基因组计划正在顺利进行之时，没多久就在人类基因组中发现了互补序列，并于 1997 年克隆了人类的端粒酶反转录酶基因 *hTERT*。

结构分析表明，TERT 蛋白的 C 端部分具有反转录酶活性，N 端部分负责与 RNA 的结合。事实上，长度为几百核苷酸的 RNA 与蛋白质结合会增加活性，从而履行其模板功能，在酶的活性位点处将核苷酸添加到正在生长的端粒上。

直到 2003 年才了解到，包括人类在内的高等真核生物的体细胞内无端粒酶活性，而生殖细胞则保留有端粒酶活性。William Hahn 及同事研究表明，培养的正常人类细胞也只是在 DNA 复制的 S 期内以较低的水平短时表达端粒酶。此外，癌细胞却具有很高的端粒酶活性且是组成型表达的。这些发现为了解癌细胞特性，甚至控制癌细胞都具有深远的意义（见知识窗 21.1）。

小结 真核生物染色体末端具有特殊的端粒结构，端粒的一条链是由许多串联重复且富含 G 的短序列组成，该重复序列具有种属特异性。富含 G 的端粒链由端粒酶合成，端粒酶的 RNA 成分是端粒合成的模板。富含 C 的端粒链通过常规的 RNA 引发 DAN 复制的方式进行合成，类似于常规后随链的合成。该机制能够确保染色体末端可被重建，避免因每轮复制而导致末端短缩。

知识窗 21.1

端粒、Hayflick 极限与癌症

每个人都知道，生物体包括人类在内都是会死亡的。但生物学家过去曾认为来源于人体的培养细胞是永生的。培养的单细胞最终会死亡，但培养的细胞系则能无限分裂。在 20 世纪 60 年代，Leonard hayflick 发现体外培养正常人类体细胞并不是永生的，在培养基上可传代数是有限的，大约为 50 代，然后进入衰老期，最终全部死亡。体外培养的正常细胞最大增殖次数称为 Hayflick 极限，但癌细胞属于永生细胞，可以无限传代。

研究发现，正常细胞与癌细胞存在很大的差异，而这种差异也许能够阐释癌细胞永生的原因。人类癌细胞内存在大量的端粒酶且是组成型表达的，而体细胞内一般只有微弱的端粒酶且是瞬时表达的（生殖细胞内必须有端粒酶的存在，以保护传给下一代的染色体的末端）。因此，癌细胞能够修复因复制而短缩的端粒，但绝大多数的正常细胞则不能修复短缩的染色体末端。所以，癌细胞能够持续分裂且不会引起染色体的降解，但正常细胞的染色体会随着细胞的每次分裂而逐渐变短，迟早会导致端粒的丢失，缺失端粒的染色体末端如同断裂染色体的末端，大多数细胞通过停止复制，最终是通过死亡来防止这种伤害，这种情况不会在癌细胞中发生，因为端粒酶能够使癌细胞免于死亡。

癌变细胞的本质性变化之一就是端粒酶基因的再激活，这是癌细胞永生化特征的前提条件，同时这也暗示了癌症潜在的治疗方法：关闭端粒酶基因的表达，或更为简便的方法是用药物抑制端粒酶活性，这种药物对大多数正常细胞是无害的，因为正常细胞一开始就无端粒酶活性。癌症研究者正致力于这一策略的研究。但是，2003 年研究发现，在培养条件下人的纤维原细胞能够低水平表达 *hTERT* 基因，纤维原细胞获得了少量的端粒酶活性，从而使这种治疗癌症的策略受到了怀疑。而且，无活性形式的 *hTERT* 基因表达或通过 RNAi 抑制正常 *hTERT* 基因表达会引起人纤维原细胞提前衰老，所以，治疗癌症的策略是既要能杀死癌细胞，又不会使患者的正常细胞提前死亡。

一些迹象表明，简单地抑制端粒酶活性可能不会导致癌细胞的死亡。首先，缺失端粒酶活性的敲除小鼠可以正常存活，尽管最终会由于端粒的缺失而导致不育，但至少能够繁殖 6 代。然而，来源于端粒酶敲除小鼠的细胞能够获得永生化特征，这些细胞能够被致肿瘤病毒转化，转化细胞移植到免疫缺陷小鼠体内后会引发肿瘤。因此，端粒酶的存在并不是癌细胞发生的绝对必要条件，或许小鼠细胞有不需要端粒酶而保存端粒的方式。我们必须要了解人类细胞是否采取不同的方式。

最后，在培养条件下人类细胞的永生使人们开始思索长生不老的问题。重新激活人的体细胞端粒酶活性能否延长寿命？或者使我们更容易患癌症？为回答这个问题，Serge Lichtsteiner、Woodring Wright 及同事将 *hTERT* 基因导入培养的人类体细胞中，使这些细胞获得端粒酶活性，实验结果令人惊奇：这些细胞的端粒被延长了，细胞在超过其正常寿命期限后仍然能够分裂，其外观及染色体含量依然如同分裂旺盛时期的细胞。而且，这些细胞并没有表现出癌变的迹象。这些发现确实是令人振奋，但并不是永葆青春的前兆，到目前为止，这还只是科学幻想。

端粒结构 除保护染色体末端、防止其降解外，端粒还有另外一个重要的功能：防止 DNA 修复系统错误地将染色体末端识别为染色体断口而将染色体粘接起来。这种异常的染色体连接可能是细胞的潜在致死因子。而且，细胞的 DNA 损伤检测系统能够检出损伤，使细胞停止分裂直至损伤修复。由于末端没有端粒的染色体看起来更像是断裂的染色体，导致 DNA 损伤检测系统启动，细胞停止分裂，最终死亡。如果端粒结构确实像图 21.29 和图 21.31 所示的那样，与染色体断口几乎没有什么差别，那么细胞如何识别端粒所具有的细微差异？

多年来，分子生物学家一直在思索这个问题，当端粒结合蛋白被发现后，他们推论，正是这些结合在染色体末端的蛋白质赋予染色体以末端特征。

在纤毛虫如 *Oxytricha nova* 中情况确实如此。图 21.32a 显示，*Oxytricha nova* 的端粒末端突出的单链包埋于**端粒末端结合蛋白**（telomere end-binding protein，**TEBP**）之中，该蛋白质包含 TEBPα 和 TEBPβ 2 个亚基，能够保护 DNA 免受降解或 DNA 修复酶的作用。酵母 *Saccharomyces cerevisiae* 的端粒蛋白是 Cdc13p，结合在酵母端粒末端的单链 DNA 上，吸引 Stn1p 蛋白的结合，Stn1p 又可招募 Ten1p 蛋白，这 3 个蛋白质共同保护酵母的端粒（图 21.32b）。

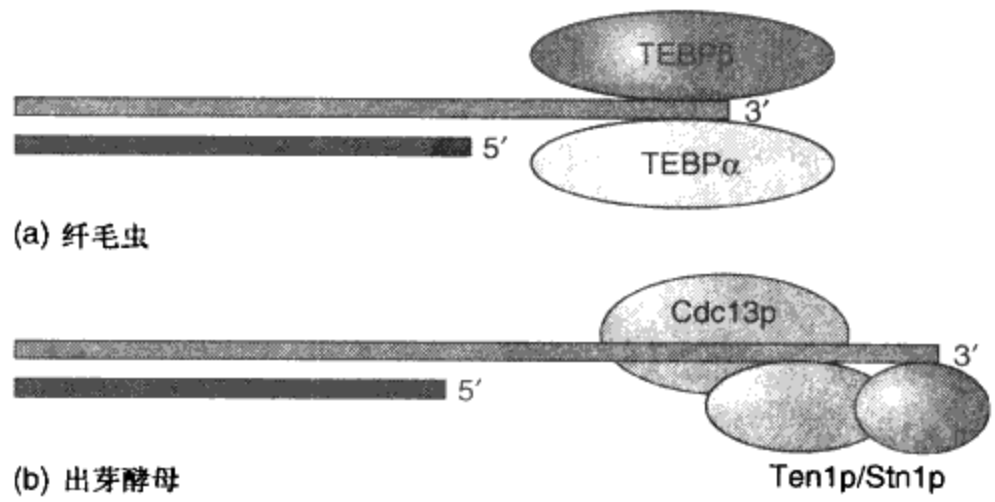


图 21.32 纤毛虫和出芽酵母的端粒结合蛋白。(a) 在纤毛虫如 *Oxytricha nova* 中，二聚体蛋白 TEBP 结合端粒末端突出的 3′ 端，起保护作用。(b) 在出芽酵母如酿酒酵母中，Cdc13p 结合在 3′ 游离端附近，招募 Stn1p 的结合，后者再招募 Ten1p 的结合。这些蛋白质在共同保护端粒末端。(*Source*: Adapted from de Lange, T., Telomere capping-one strand fits all. *Science* 2001 [2001] p. 1075, f. 1.)

有关哺乳动物和裂殖酵母的单链端粒结合蛋白我们将在本节后面讨论。然而，哺乳动物还可以通过 DNA 二级结构来保护端粒突出的单链。对这个问题的研究始于哺乳动物端粒 R 结合蛋白的发现。如 TTAGGG 重复结合因子 TRF1 和 TRF2，这两个蛋白质在端粒区能够与双链而不是单链 DNA 发生特异结合。直到 2001 年，还没有人在哺乳动物中发现能与端粒最末端（图 21.31 所示的富含 G 的单链

突出端）发生特异结合的蛋白质。然而在 1999 年，Jack Griffith 和 Titia de Lange 及同事发现哺乳动物的端粒并不是以前所认为的线性结构，而是形成 DNA 噜噗，即 **t 环**（telomere loop）。这些环状结构在染色体上是唯一的，很容易使端粒末端与因染色体中部断裂而形成的线性末端区分开来。

有什么证据能够证明 t 环的存在？Griffith、de Lange 及同事首先合成了 TTAGGG 重复序列达 2kb 的典型哺乳动物端粒 DNA，其末端为 150～200nt 的单链 3′突出端，加入端粒结合蛋白 TRF2，对形成的复合物进行电子显微镜观察。图 21.33a 显示确实存在 DNA 噜噗的形成，球形蛋白质 TRF2 恰好位于环-尾结点处，此结构出现的概率约为 20%，相反，当研究者切除单链的 3′突出端或不加入 TRF2 蛋白，形成噜噗的可能性急剧降低。

端粒形成这样环形结构的途径之一可能是，单链 3′突出端侵入了双链端粒 DNA 的上游区域，如图 21.34 所示。如果这个假设是正确的，则可用补骨脂素和 UV 照射处理，使双链 DNA 链上的胸腺嘧啶发生交联而稳定环形结构。因为侵入链能够与被侵入双链中的一条链进行碱基配对，形成可以交联的双链 DNA，从而可以稳定形成的环形结构。图 21.33b 显示了 Griffith、de Lange 及同事的实验结果。研究者用补骨脂素和 UV 照射处理，使模式 DNA 发生交联，然后将复合体进行脱蛋白质处理，于电子显微镜下观察分子结构。结果发现，即使无 TRF2 蛋白的存在，环形结构依然清晰可见，说明 DNA 分子通过自身的交联而实现环形结构的稳定。

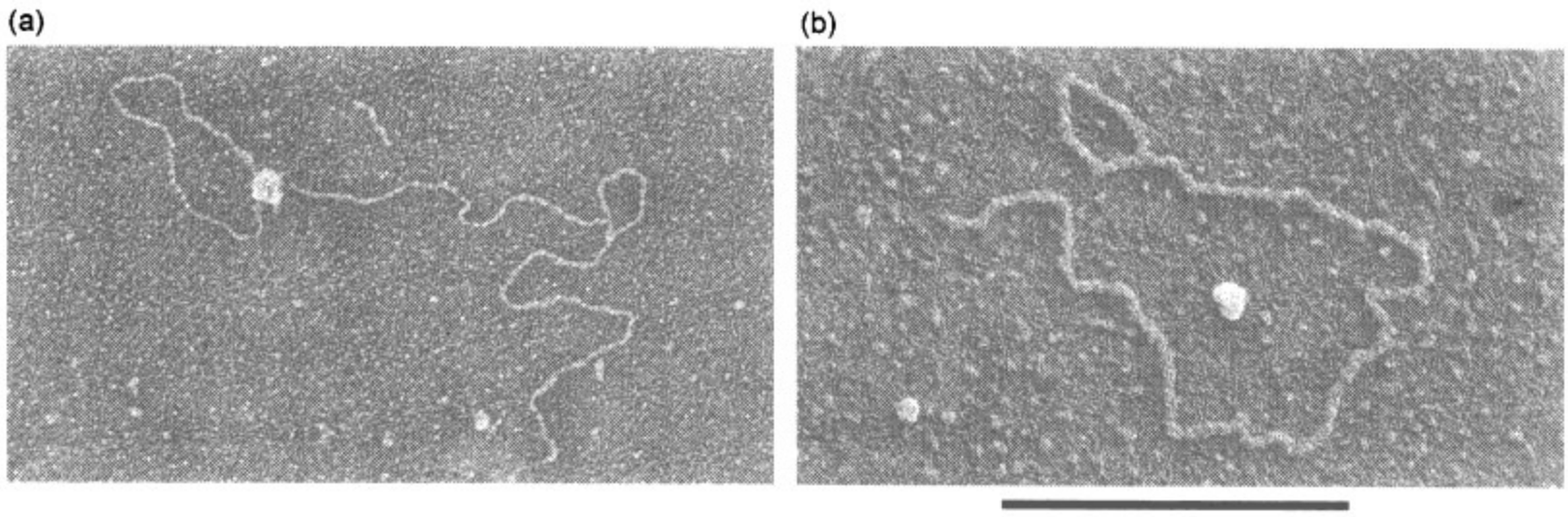


图 21.33　t 环的体外形成。（a）环的直接检测。Griffith 及同事将具有端粒类似结构的典型 DNA 与 TRF2 混合，然后将混合物涂展在 EM 载片上，用钨喷镀 DNA 和蛋白质使其形成阴影，电镜观察形成阴影的分子。图中出现一个明显的环，TRF2 蛋白位于环和尾的连接处；（b）通过交联而稳定环形结构。Griffith 及同事制备了图（a）中的 t 环，然后经补骨脂素和 UV 照射使双链 DNA 交联，去除蛋白后，将交联的 DNA 涂展在 EM 载片上，用铂和钯喷镀使之形成阴影，进行电镜观察发现，存在一个明显的环，线条表示 1kb。（*Source*：Griffith，J. D.，L. Comeau，S. Rosenfield，R. M Stansel，A. Bianchi，H. Moss and T. de Lange，Mammalian telomeres end in a large duplex loop. *Cell* 97（14 May 1999）f. 1，p. 504. Reprinted by permission of Elsevier Science.）

为进一步证实观察到的环形结构包含有端粒

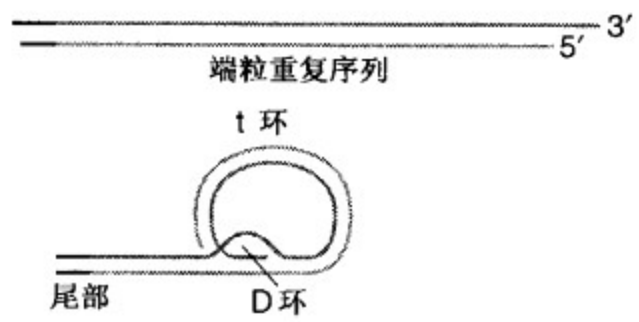


图 21.34　哺乳动物的 t 环模型。富含 G 链（红色）的单链 3′端侵入双链端粒 DNA 的上游区域，形成一个长的 t 环和环尾连接处 75～200nt 的置换环。图中显示邻近端粒（蓝色和红色）的短亚端粒区（黑色）。（*Source*：Adapted from Griffith，D.，L. Comeau，S. Rosenfield，R. M. Stansel，A. Bianchi，H. Moss and T. de Lange，Mammalian telomeres end in a large duplex loop. *Cell*，97：511，1999.）

DNA 的观点，Griffith、de Lange 及同事向环形 DNA 中加入能与端粒双链 DNA 特异结合的 TRF1 蛋白。

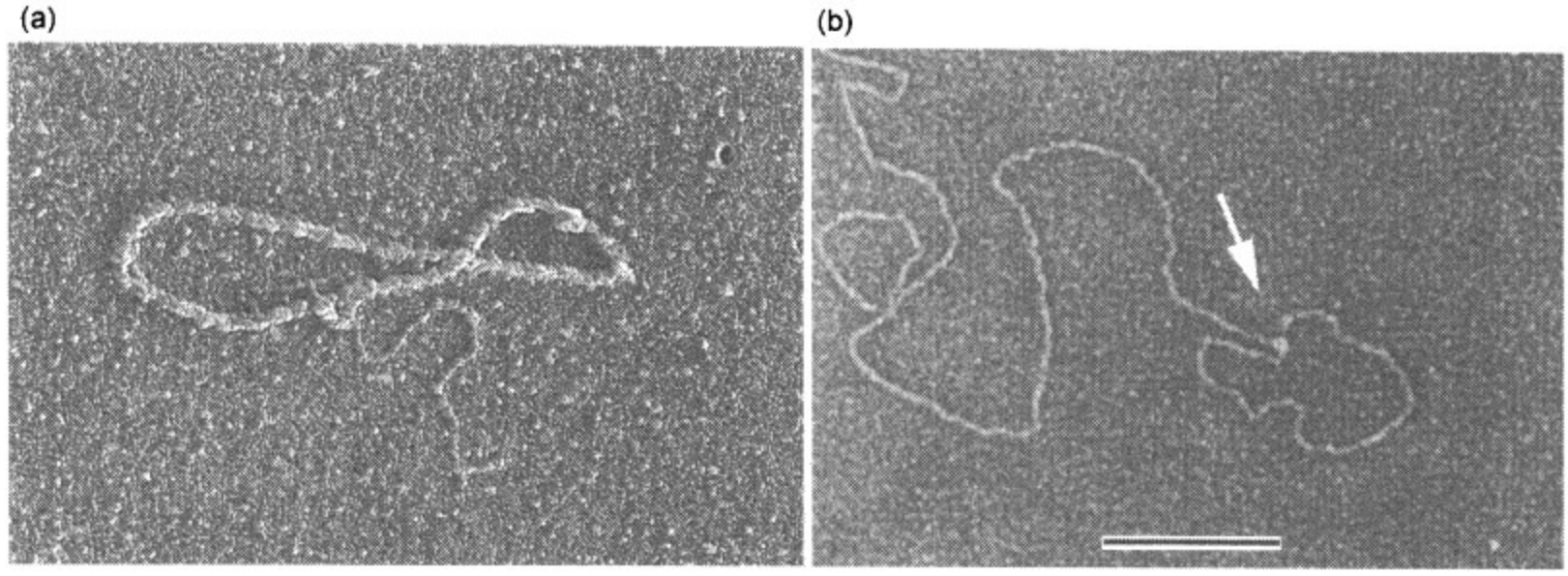


图 21.35　TRF1 和 SSB 与 t 环的结合。（a）TRF1。Griffith、de Lange 及同事纯化出天然的 HeLa 细胞 t 环，用补骨脂素和 UV 照射处理使之发生交联，然后加能与端粒双链 DNA 特异结合的 TRF1 蛋白。用铂和钯喷镀使环形成阴影后于电子显微镜下观察。t 环（而不是尾部）由 TRF1 均一地覆盖。（b）SSB。研究用*E. coli*的 SSB 代替 TRF1，进行上述实验，SSB 应该与单链 DNA 结合，确实观察该蛋白质结合在环-尾结合处（箭头所示），此处正是预计的单链取代环的位置。线条代表 1kb。（*Source*：Griffith，J. D.，L. Comeau，S. Rosenfield，R. M. Stansel，A. Bianchi，H. Moss and T. de Lange，Mammalian telomeres end in a large duplex loop. *Cell* 97（14 May 1999）f. 5，p. 510. Reprinted by permission of Elsevier Science.）

如图 21.35a 所示，研究者观察到 TRF1 蛋白覆盖在环形 DNA 上。

如果图 21.34 所示的链侵入假说是正确的，那么被侵入 DNA（置换环或 D 环）所置换出的单链 DNA 在达到足够长度后，就应该能够与*E. coli*的单链结合蛋白（SSB，参阅第 20 章）结合。图 21.35b 表明，SSB 确实存在于尾-环交接处。这正是我们预测置换 DNA 的位置。

TRF1 和 TRF2 在 t 环形成过程中有何作用？图 21.33b 表明，TRF2 存在于能够发生链侵入的尾-环交接处，而且我们也了解到，没有 TRF2 蛋白的存在就不会有环形结构的形成。所以，Griffiths、de Lange 及同事推测该蛋白在链侵入过程中起作用，并与链侵入位点结合。图 21.35a 显示 TRF1 蛋白与 t 环中的双链端粒 DNA 结合，而且，TRF1 能够诱导双链 DNA 发生弯曲，并与双链端粒 DNA 结合，所以，Griffiths、de Lange 及同事推测该蛋白质具有弯曲双链端粒、维持环形构象的作用，从而利于链的入侵。图 21.36 对这些推论进行了总结。DNA 弯曲的最终结果是使敏感的染色体末端隐藏于置换环内，避免被 DNA 修复酶识别为染色体的断口，并将其与另一染色体末端粘接，进而造成严重后果。

2001 年，Peter Baumann 和 Thomas Cech 报道在裂殖酵母（*Schizosaharomyces pombe*）和人等生物体内发现了能与端粒单链尾部结合的蛋白质。研究者在酵母和人类基因组数据库中搜索同 TEBPα 的一部

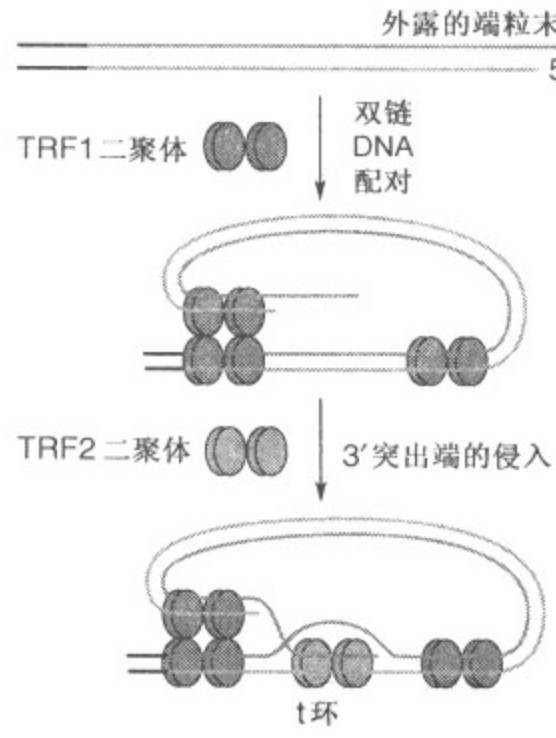


图 21.36　TRF1 和 TRF2 在哺乳动物 t 环形成中的作用。TRF1（绿色）通过与端粒的双链 DNA 部分结合而使 DNA 发生弯曲，使之保持利于链侵入的构象。TRF2（橙黄色）也能与 DNA 双链区结合，但它的主要作用可能是结合在链的侵入位点，稳定置换环，进而稳定 t 环。（*Source*：Adapted from Griffith，J. D.，L. Comeau，S. Rosenfield，R. M. Stansel，A. Bianchi，H. Moss and T. de Lange，Mammalian telomeres end in a large duplex loop. *Cell*，97：511，1999.）

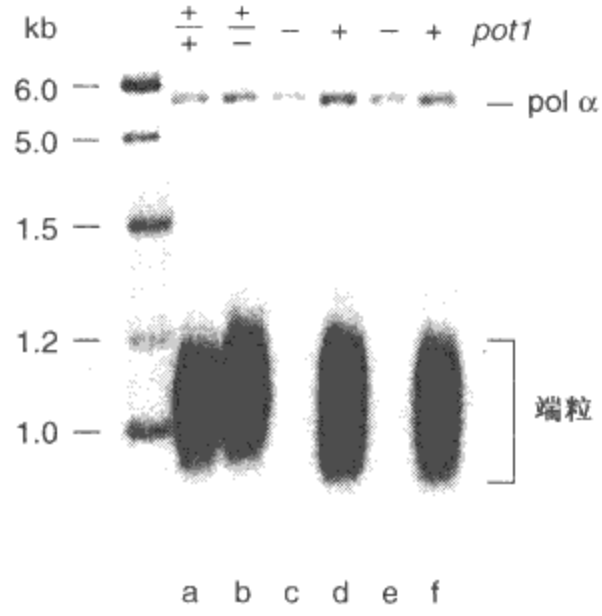


图 21.37　*pot1* 基因缺陷的裂殖酵母菌株的端粒丢失。Baumann 和 Cech 制备了 *S. pombe* 的纯化与杂合双倍体，以及 $pot1^-$ 与 $pot1^+$ 单倍体菌株，如图上方所示。然后分离这些菌株的 DNA，用 *Eco*RI 酶切，电泳后，用端粒特异性探针进行 Southern 斑点杂交。作为与均匀加载斑点的对照，对 DNA 聚合酶 α 也进行检测，如图右上方所示。（*Source*: From Baumann and Cech, *Science* 292: p. 1172. © 2001 by the AAAS.）

分相匹配的序列，结果发现，这两个物种的序列相似性有限，相似程度最高的序列位于蛋白质的氨基末端，长度约为 20 个氨基酸。该序列在纤毛虫中也是最为保守的。因为 TEBP α 是纤毛虫的端粒保护蛋白，Baumann 和 Cech 将 *S. pombe* 内的同源基因命名为 *pot1*（*protection of telomeres 1*）。

为验证 *pot1* 编码端粒保护蛋白质的假设，Baumann 和 Cech 使 $pot1^+/pot1^-$ 二倍体菌株形成孢子，与 $pot1^+$ 型孢子相比，$pot1^-$ 型孢子形成的克隆很小，$pot1^-$ 细胞呈长形，存在染色体分离缺陷从而导致细胞停止分裂。所有这些效应都与端粒功能丧失相一致。为检测 *pot1* 对端粒的直接效应，Baumann 和 Cech 用端粒特异性探针对来自 $pot1^-$ 菌株的 DNA 进行 Southern 斑点杂交，筛选具有端粒结构的细胞。图 21.37 显示了实验的结果。来自 $pot1^+$ 菌株的 DNA 以及来自至少含有一个 $pot1^+$ 等位基因的二倍体菌株的 DNA 都能够与端粒探针发生强烈反应，说明这些菌株含有端粒。但是，来自 $pot1^-$ 菌株的 DNA 不能与端粒探针发生反应，表明这些菌株的端粒已经消失。所以，*pot1* 似乎确实具有保护端粒的作用。

pot1 基因的蛋白质产物是 Pot1p，如果该蛋白确实能够保护端粒，那么我们会认为此蛋白质会结合在端粒上。为验证我们的预测，Baumann 和 Cech 将 *pot1* 基因克隆到 *E. coli* 载体，表达后能够产生具有 6 个组氨酸作为标签（参见第 4 章）的融合蛋白。然后对纯化的融合蛋白进行凝胶迁移率分析（参见第 5 章），从而鉴定该蛋白质是与富 C 的端粒链还是与富 G 的端粒链，或者是与端粒双链 DNA 结合。图 21.38a 显示，Pot1p 蛋白是与富 G 的端粒链结合，而不与富 C 的端粒链以及端粒双链 DNA 结合，而且，Pot1p 蛋白的 N 端片段更能高效地与富 G 端粒链结合（图 21.38b）。

随后，Baumann 和 Cech 在人类蛋白质数据库中进行 BLAST 搜索（参见第 24 章），发现与 Pot1p 相似的蛋白质 hPot1p。hPot1p 蛋白与 *S. pombe* 的 Pot1p 蛋白在高度保守的氨基末端区域具有 26%的一致性和 48%相似性。研究者克隆了目的基因并使之表达产生 His_6-融合蛋白，其 C 端片段是不完全的。hPot1p 的 N 端片段与 *S. pombe* 蛋白的性质一样，都能结合端粒 DNA 富 G 链（图 21.38c）。但是，hPot1p 的 N 端片段只能与人的端粒 DNA 结合。进一步的同型实验结果表明，*S. pombe* 和 human 蛋白只能与自身端粒 DNA 结合，不能够彼此相互结合。

单链端粒结合蛋白 hPot1p 与隐藏单链端粒 DNA 的 t 环是如何协同存在的？Baumann 和 Cech 推测，哺乳动物的端粒以 3 种不同的状态存在：t 环、hPot1p 结合态和端粒酶结合态。这 3 种状态可能都会对端粒起到保护作用，防止端粒的损毁以及遭到双链断裂修复系统的连接。单链 DNA 的最末端也可能并没有隐藏在 t 环中，此时需要 hPot1p 蛋白起到“保护帽”的作用。

$pot1^-$ 菌株有一个非常有趣的现象，尽管最初是十分异常的，但在繁殖大约 75 代后即可转化正常菌株。这种效应也曾在端粒酶缺失的菌株中观察到。如果酵母染色体的端粒缺失，则可以通过染色体的环化来保护其末端。为验证这个假设，Baumann 和 Cech 用稀有切点限制性内切核酸酶 *Not*I（参见第 4 章）切割 $pot1^-$ 菌株的 DNA，对产生的 DNA 片段进行脉冲电场凝胶电泳分析，如果染色体确实发生环化，染色体末端的 *Not*I 片段就会消失，取而代之的是末端融合后形成的新片段。图 21.39 显示，被检测的染色体 I 和 II 完全发生了上述情况。染色体 I 末端正常的 2 个 *Not*I 片段（I 和 L）消失，而产生了新片段（I+L），$pot1^+$ 菌株无此现象的出现。同样，染色体 II 末端正常的 2 个 *Not*I 片段（C 和 M）消失，而产生了新片段（C+M）。所以，$pot1^-$ 菌株的染色体确实因端粒的丢失而发生了环化。

小结　所有真核生物都能保护其端粒结构免受核酸酶及双链断裂修复酶的作用。纤毛虫 TEBP 蛋白的结合可保护端粒 3′突出的单链末端，

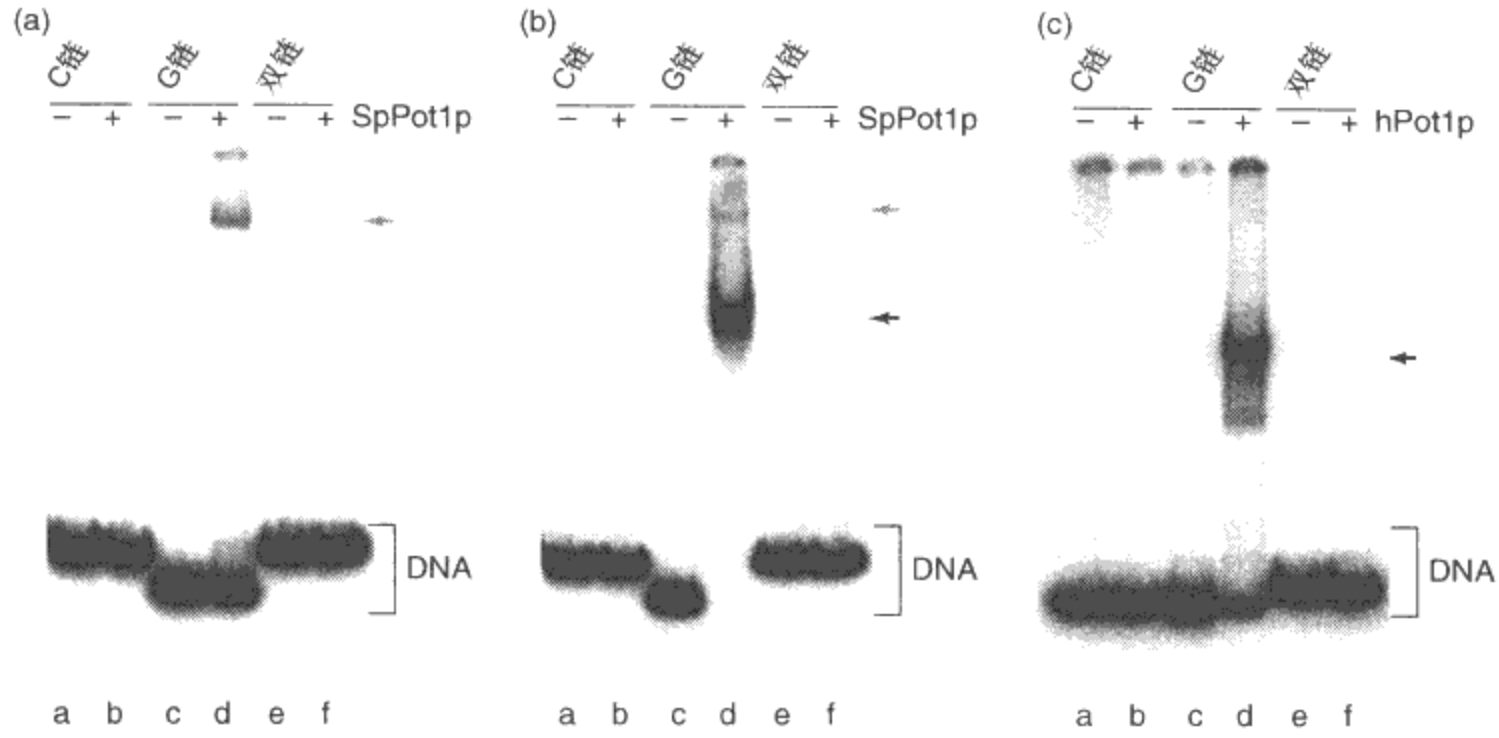


图 21.38　Pot1p 与端粒 DNA 的结合。Baumann 和 Cech 对 *S. pombe* 的 Pot1p 蛋白和标记的 *S. pombe* 端粒 DNA（a 和 b）及人的 hPot1 蛋白和标记的人的端粒 DNA（c）进行凝胶迁移率分析。端粒 DNA 的富 C 链、富 G 链或双链 DNA 标注在图上方。图（a）含有全长 Pot1p。图（b）含有 Pot1p N 端的大部分片段，并有少量全长 Pot1p 蛋白污染。图（c）含有 hPot1p N 端片段。箭头指示含有全长 Pot1p（黄色）或含有 Pot1p N 端的部分片段及 Pot1p 或 hPot1p N 端（蓝色）变动后条带的位置。（*Source*：From Baumann and Cech, *Science* 292：p. 1172. © 2001 by the AAAS.）

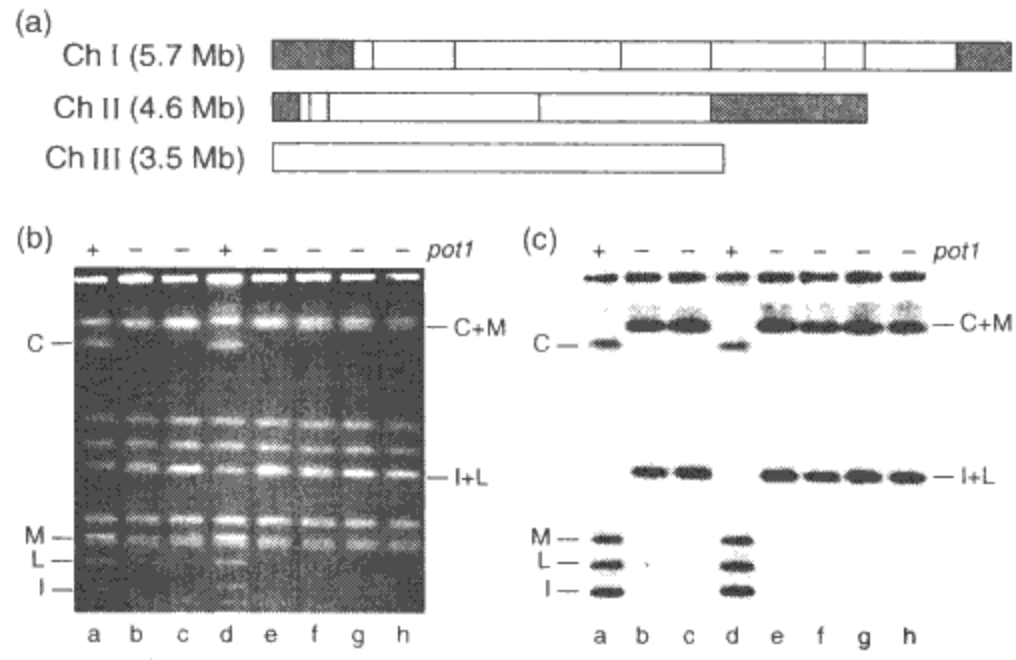


图 21.39　存活的 *pot1*⁻菌株含有环化的染色体。（a）*S. pombe* 3 条染色体的 *Not*I 酶切位点图谱，竖线表示酶切位点。染色体 I 和 II 的 *Not*I 末端片段用红色表示。染色体 III 无酶切位点，所以不能被 *Not*I 切割。（b）对来自 *pot1*⁻和 *pot1*⁺细胞（标注于图上方）的 *Not*I DNA 片段进行脉冲电场凝胶电泳，然后将凝胶染色。染色体 I 和 II 末端片段（C、M、L 和 I）的位置在左边标出，融合片段 C+M 和 I+L 的位置在右边标出。（c）Baumann 和 Cech 用分别代表染色体 I 和 II 末端片段的标记 DNA 片段 C、M、L 和 I 为探针，对图（b）的凝胶进行 Southern 斑点检测，显示染色体 I 和 II 的末端。

出芽酵母蛋白 Cdc13p 能够招募 Stn1p 和 Ten1p 并与单链端粒 DNA 结合。哺乳动物和裂殖酵母也有类似于 TEBP 的蛋白，能够与单链端粒 DNA 发生特异性结合。此外，哺乳动物的端粒可以形成环状结构以保护单链端粒 DNA。在哺乳动物的细胞内，TRF1 和 TRF2 显然能够帮助端粒 DNA 形成 t 环，在 t 环中，端粒的 3′单链末端侵入到双链端粒 DNA 上游。TRF1 能够使 DNA 发生弯曲，有助于链的侵入，而 TRF2 与链侵入点结合，稳定置换环。

总结

在 *E. coli* 中，引物的合成需要引发体的参与，引发体由 DNA 解旋酶、DnaB 及引发酶 DnaG 组成。引发体在复制原点 *oriC* 处的组装过程如下：DnaA 与 *ori C* 内的 *dnaA* 盒位点结合，与 RNA 聚合酶、HU 蛋白协同使邻近 *dnaA* 盒最左侧的 DNA 区域发生解聚，然后 DnaB 与开放复合体结合，并促使引发酶的结合，从而完成引发体的组装。引发体与复制体结合在一起，重复引发后随链冈崎片段的合成。DnaB 同时具有旋转酶活性，能够在复制体移动过程中解链 DNA。

SV40 的复制原点邻近病毒转录调控区，复制起始于大 T 抗原与 *oriC* 位点的结合。大 T 抗原与 64bp 的 *ori* 核心及相邻的 2 个位点结合，行使解旋酶活性将 DNA 解旋，在 *ori* 核心内产生复制泡，由与真核生物 DNA 聚合酶 α 有关的引发酶引发 SV40 DNA 的复制。

酵母的复制原点位于由 4 个重要区域（A、B1、B2 和 B3）组成的自主复制序列（ARSs）内。A 区长 15bp，其中的 11bp 序列在 ARSs 内是高度保守的；B3 区会使 ARSs 内的 DNA 形成重要的弯曲。

在体外，pol III 全酶能以 730nt/s 的速率合成 DNA，稍慢于体内 1000nt/s 的合成速率，无论在体内或体外，pol III 全酶均具有很高的进行性。

pol III 核心聚合酶（αε 或 αεθ）酶自身无进行性功能，在脱离模板之前只能复制一段短的 DNA 片段。相比之下，核心酶与 β 亚基结合后会以 1000nt/s的合成速率进行 DNA 复制。β 亚基二聚体形成环形结构，环绕 DNA 模板，并与 α 亚基相互作用，将模板和聚合酶结合在一起，使聚合酶能够长时间地停留在模板上而具有很高的进行性。真核生物的进行性因子 PCNA 蛋白形成环形三聚体结构，通过环绕 DNA 分子而将聚合酶结合在模板上。

β 亚基向 DNA 复合体上装载需要 γ 复合体（γ、δ、δ′、χ 和 Ψ）的协助，γ 复合体对进行性 αδβ 复合体的形成起催化作用。在进行性的复制过程中，γ 复合体并不与 αδβ 复合体保持结合。β 钳的装载是一个依赖 ATP 的过程。

pol III 全酶的 2 个核心聚合酶通过 2 个 τ 亚基与 γ 复合体连接在一起。其中一个核心聚合酶负责前导链的连续合成，另一个核心聚合酶负责后随链的不连续合成。γ 复合体作为钳装载器将 β 钳装载到具引物的 DNA 模板上。一旦完成装载，β 钳失去与 γ 复合体的亲和力，转而与核心聚合酶结合，帮助核心聚合酶高度进行性地合成冈崎片段。一旦完成冈崎片段的合成，β 钳便失去与核心聚合酶的亲和力，进而再与 γ 复合体结合，γ 复合体作为钳卸载器将 β 钳从 DNA 模板上释放下来，使之再循环到下一个引物处，起始新一轮的复制，实现此过程的重复。

在复制过程的末尾，环状细菌染色体形成了需通过两步过程才能解离的环连体。首先，连接两条链未被复制的双螺旋解聚，然后修复合成填补空隙。形成了具有偶数节点的右手平行螺绕环连体，由拓扑异构酶 IV 从节点处将环连体解开。线性真核生物染色体在 DNA 复制过程中也需要解连接。

真核生物染色体末端具有特殊的端粒结构，端粒的一条链是由许多串联重复且富含 G 的短序列组成，该重复序列具有种属特异性。富含 G 的端粒链由端粒酶合成，端粒酶的短 RNA 分子是端粒合成的模板。富含 C 的端粒链通过常规的 RNA 引发 DAN 复制的方式进行合成，类似于常规后随链的合成。该机制能够确保染色体末端可被重建，避免因每轮复制而导致末端短缩。

所有真核生物都能够保护其端粒结构，避免受到核酸酶及双链断裂修复酶的作用。纤毛虫 TEBP 蛋白的结合能够保护端粒 3′-突出的单链末端，出芽酵母蛋白 Cdc13p 负责招募 Stn1p 和 Ten1p 并与单链端粒 DNA 结合。哺乳动物和裂殖酵母也有类似于 TEBP 的蛋白，能够与单链端粒 DNA 发生特异性结合。此外，哺乳动物的端粒可以形成环状结构以保护单链端粒 DNA。在哺乳动物的细胞内，TRF1 和 TRF2 显然能够帮助端粒 DNA 形成 t 环，在 t 环中，端粒的 3′-单链末端侵入到双链端粒 DNA 上游。TRF1 能够使 DNA 发生弯曲，有助于链的侵入，而 TRF2 与链侵入点结合，稳定置换环。t 环结构能够保护染色体末端，防止发生将两个染色体末端粘接在一起的错误修复行为。

复习题

1. 叙述定位并确定最短复制原点的实验。

2. 列举 *E. coli* 引发体的组分及其在引物合成中的作用。

3. 简要叙述定位 SV40 DNA 复制原点的策略。

4. 简要叙述鉴定酵母自主复制序列（ARS1）的策略。

5. 简要叙述证明酵母 DNA 复制始于 ARS1 的策略。

6. 叙述实验并给出结果，显示体外 DNA 链的延伸速率。

7. 叙述体外检测 DNA 合成进行性的实验程序。

8. pol III 全酶中的哪个亚基负责进行性？哪个亚基负责将进行性亚基（滑动钳）装载到 DNA 上？

该滑动钳与哪个核心亚基结合?

9. 叙述实验并给出结果，说明β钳在环状DNA和线性DNA上具有不同行为，这种不同行为对研究β钳与DNA之间的相互作用模型有什么提示?

10. X射线晶体衍射图像的研究暗示了β钳与DNA之间怎样的相互作用模型?

11. X射线晶体衍射图像的研究暗示了PCNA与DNA之间怎样的相互作用模型?

12. 叙述实验并给出结果，证明钳装载器具有催化功能，钳装载器的成分是什么?

13. 简要概述解释钳装载器如何利用ATP能将β钳打开并让DAN进入的假说。

14. 不连续合成的后随链如何与连续合成的前导链保持同步?

15. 描述证明 pol III* 可与β钳解聚的实验方案，并阐述实验结果。

16. 描述蛋白质足迹技术的实验程序。说明该技术被用于证明 pol III 核心酶与滑动钳装载器均与β钳同一位点发生相互作用的实验步骤。

17. 叙述实验并给出结果，说明γ复合体具有钳状卸载器功能。

18. 描述在DNA不连续复制中，β钳是如何通过与 pol III 核心酶及钳卸载器之间的结合而实现循环?

19. 环形DNA复制结束后，为什么需要解连环?

20. 叙述实验并给出结果，证明拓扑异构酶IV是 *Salmonella typhimurium* 和 *E. coli* 质粒解连环所必需的。

21. 为什么真核生物需要端粒，而原核生物不需要?

22. 图示说明端粒合成过程。

23. 为什么说四膜虫是研究端粒的最好模式生物?

24. 描述分析端粒酶活性的实验，简要说明实验结果。

25. 叙述实验并给出结果，证明端粒酶RNA是端粒合成的模板。

26. 图示说明端粒的t环结构。

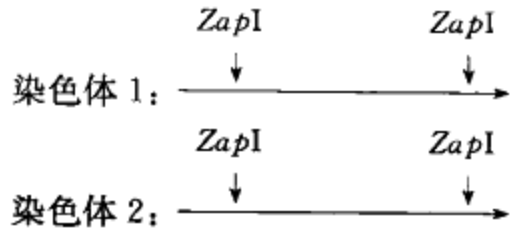
27. t环存在的实验证据是什么?

28. t环形成过程中，支持链入侵的实验证据是什么?

分析题

1. 从人类 *hpot1* 基因的核苷酸序列或hPot1蛋白的氨基酸序列开始，如何在另一基因组已知生物如线虫 *Caenorhabditis elegans* 中寻找同源基因或同源蛋白? 如何获得该蛋白质并验证其Pot1蛋白活性?

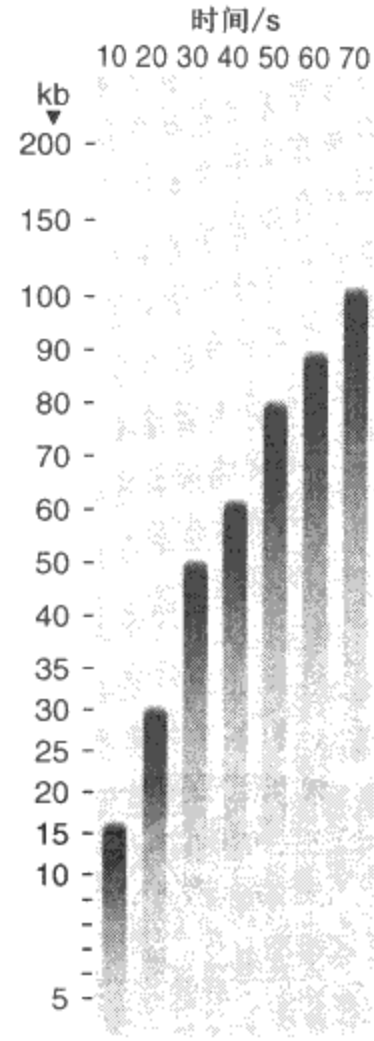
2. 你正在研究一新发现的原生动物的 *pot1* 基因，发现 *pot1* 缺陷细胞在繁殖50代后回复到正常细胞状态，野生型细胞只有2条染色体，其限制性内切核酸酶 *Zap*I 的酶切图谱如下：



提出突变细胞回复到正常细胞的假设，描述证明该假设的实验方案及预期结果。

3. 你正在研究一真核生物病毒的双链DNA基因组，其长度为130kb。推测在基因组上可能有多个复制原点，设计实验验证你的假设，并找出所有的复制原点。

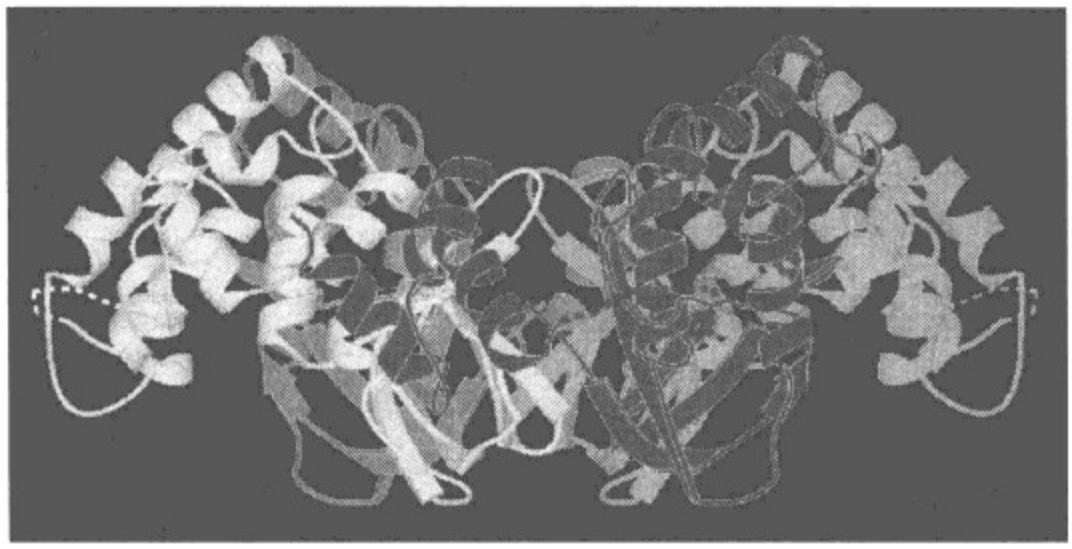
4. 你正在研究一新发现的极端嗜热细菌 *Rapidus royi* DNA复制的延伸速率。将不同反应时间获得的DNA延伸产物进行电泳，结果如图所示，延伸速率如何? 是否为生物界中最高的延伸速率?



5. 你正在研究一个新种细菌的DNA复制，发现该生物具有与 *E. coli* 相类似的β钳和 pol III*，你想知道在DNA非复制期间及模式DNA模板复制终止后β钳与 pol III* 是否发生分离。叙述实验方案并给出预期结果，包括分离分析。

翻译　徐启江　校译　李玉花

第22章 同源重组



RuvA 四聚体的晶体结构，每个单体以不同颜色表示。在 *E. coli* 的重组过程中，RuvA 可与 Holliday 结构结合，并促进 Holliday 结构的分支端化。(*Source: Rafferty, J. B., S. E. Sedelnikova, D. Hargreaves, P. J. Artymink, P. J. Baker, G. J. Sharples, A. A. Mahdi, R. G. Lloyd and D. W. Rice, Crystal structure of DNA recombination protein RuvA and a model for its binding to the Holliday junction. Science 274 (18 Oct 1996) f. 2e, p. 417. Copyright © AAAS.*)

遗传学家早已知道，通过有性生殖繁衍的后代往往具有与父母双亲不同的遗传组成。这种变异除了源自减数分裂过程中双亲染色体的自由组合外，绝大部分来源于同源染色体之间的同源重组。这个过程改变了父源和母源染色体上基因的组成，使后代个体中出现非亲本型的重组染色体。这些新组合使得后代个体能获得比亲本更具优势的存活机会，这种重组型是有意义的。另外，在减数分裂过程中，同源染色体间的重组所形成的物理连接保证了前期同源染色体间的正确配对和中期的完全分离。正如在第 20 章中所论述的，同源重组在细胞 DNA 修复过程中也起到了非常重要的作用，这一过程被称作重组修复。

图 22.1 描述了同源重组过程中的几种不同事件。而对每一事件都是通过一种“将事先分开的 DNA 片段连接进而实施交换”的过程加以阐明的。但这并不表明交换仅起始于两个彼此独立的 DNA 分子之间。交换也可发生在 DNA 分子内，这种分子内的交换发生在同一染色体的两个不同位点之间，它将导致这两个位点之间的序列丢失或倒位。此外，两分子间的重组涉及两个独立 DNA 分子之间的交换。重组通常是相互的，即参与重组的两个 DNA 片段是双向互换的。DNA 分子可以发生一次、二次或多次交换事件，交换的次数明显地影响着最终产物的特性。

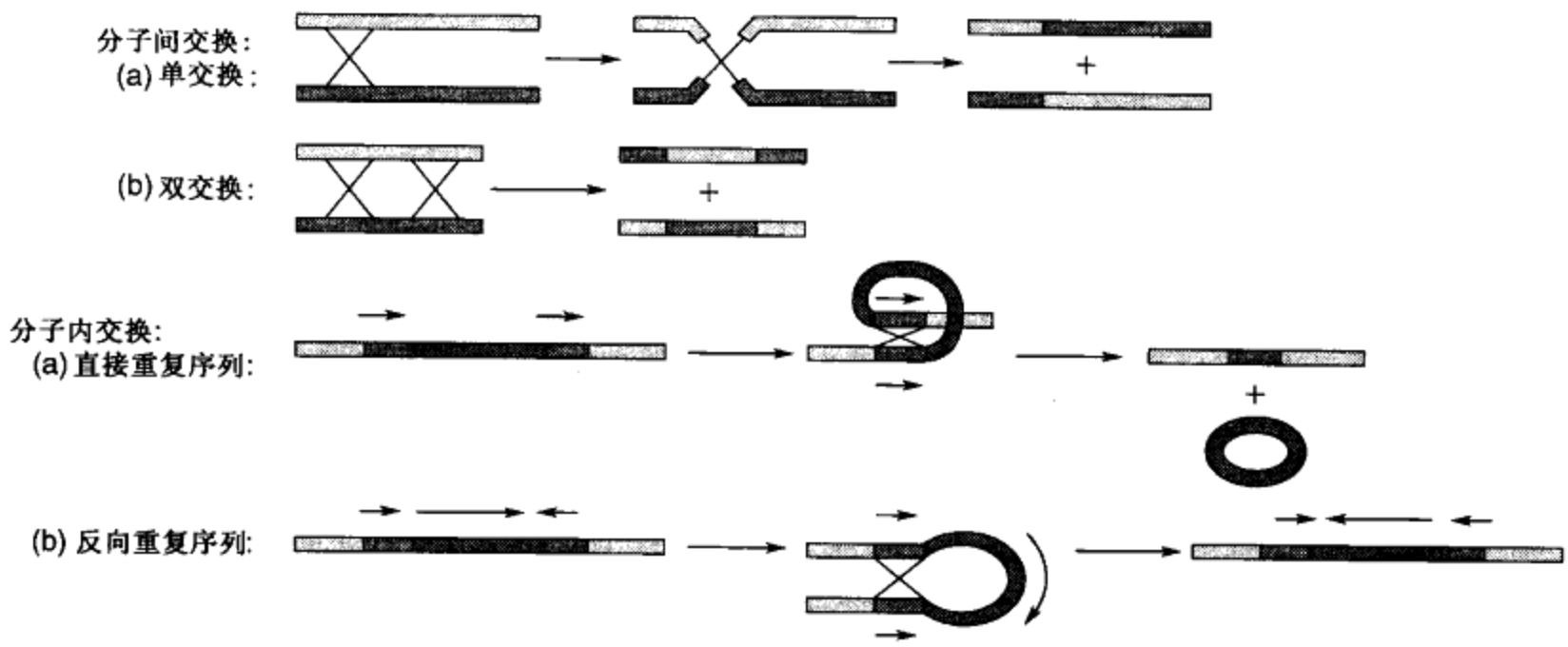


图 22.1　重组实例。图中“×”代表两条染色体或者同一染色体不同部分之间发生的交换事件。从上图所示的相互重组中间体的形成可以了解重组的过程。可以想象这一 DNA 链断裂后形成新的 DNA 交叉连接（用“×”表示，下同）。

22.1　同源重组的 RecBCD 途径

为了阐明同源重组的原理，我们将讨论已在 *E. coli*中被研究得很透彻的 RecBCD 途径。

该重组过程（图 22.2）起始于对重组 DNA 分子中的一条 DNA 进行双链切割的诱导。RecBCD 蛋白（*RecB*、*RecC* 和 *RecD* 基因的产物）结合到 DNA 双链断裂区，并利用其 DNA 解旋酶活性向具有 5′-GCTGGGTGG-3′序列的 *Chi* 位点进行 DNA 解旋。在 *E. coli*基因组中平均每 5000bp 就有一个 *Chi* 位点。RecBCD 蛋白具有外切双链和单链的活性，也具有单链内切核酸酶活性。这就使得 RecBCD 蛋白可以产生单链末端，然后这个单链末端可以被 RecA 蛋白（*recA* 基因的产物）和单链 DNA 结合蛋白（SSB）包裹。RecBCD 蛋白也协助 RecA 装载到 3′-DNA 末端。

RecA 蛋白使这个单链末端侵入其他双链 DNA 寻找同源区域。这一侵入将产生一个置换环（D 环）——因 DNA 链的置换（displaced）而得名。一旦该 DNA 末端发现同源区域，DNA 的 D 环区域会在 RceBCD 蛋白的帮助下形成一个缺口。在该缺口处，RecA 蛋白和 SSB 能产生一个与另一条 DNA 的间隙区域互补配对的新末端。DNA 连接酶催化封闭缺口，并产生一个**Holliday 联接体**（Holliday junction），以 1964 年第一个提出这个假设的学者 Robin Holliday 的名字命名。Holliday 联接体也称为半交叉结（half chiasma）和 *Chi* 结构。在 Holliday 联接体中的分支行为可以将已经断裂的碱基对形成新的碱基对并向任一方向移动，这一过程也称为分支端化（branch migration）。

自发产生分支端化的频率很低。就像在 DNA 复制过程中需要解旋酶和来自 ATP 的能量一样。分支端化过程中，RuvA 和 RuvB 两个蛋白质协作来行使这些功能。它们都具有 DNA 解旋酶活性，而且 RuvB 是一种 ATP 酶，能将储存在 ATP 中的能量用于分支端化过程。最后，两条 DNA 链必须形成缺口以拆分 Holliday 联接体，进而形成异源双链或重组产物，RuvC 蛋白行使将 Holliday 联接体拆分的功能。

RecBCD 蛋白是怎样产生具有 3′端的单链 DNA 并起始链侵入的呢？图 22.3 阐明了两种依赖于 ATP 和 Mg^{2+} 相对浓度的可能机制。

当 ATP 浓度大于 Mg^{2+} 浓度时，大部分的 Mg^{2+} 被螯合。当体外实验处于这种状态时，依照图 22.3 左面的途径进行。RecBCD 蛋白的解旋酶活性向着 *Chi* 位点方向打开 DNA 双链，形成一个具有 3′端的环结构。一旦 RecBCD 到达 *Chi* 位点，就会切割 3′端链上 *Chi* 位点下游的核苷酸，并且 RecA 蛋白包裹环突部分的 DNA。RecBCD 将继续打开 *Chi* 位点上游的 DNA 螺旋，但不再进一步降解 DNA 了。

此外，当 Mg^{2+} 浓度大于 ATP 浓度时，会有大量游离的 Mg^{2+}。在体外实验处于这种状态时，按照图 22.3 右图的途径进行。同样，RecBCD 的解旋酶活性向着 *Chi* 位点方向解旋 DNA 双链，形成一个具有 3′端的环结构。但此时 RecBCD 降解环下游的 3′端，当 RecBCD 到达 *Chi* 位点时，由于 RecA 蛋白结合到环突 DNA 环上，停止降解 3′端。而当 RecBCD 越过 *Chi* 位点继续解开双链 DNA 时将降解 5′端链。

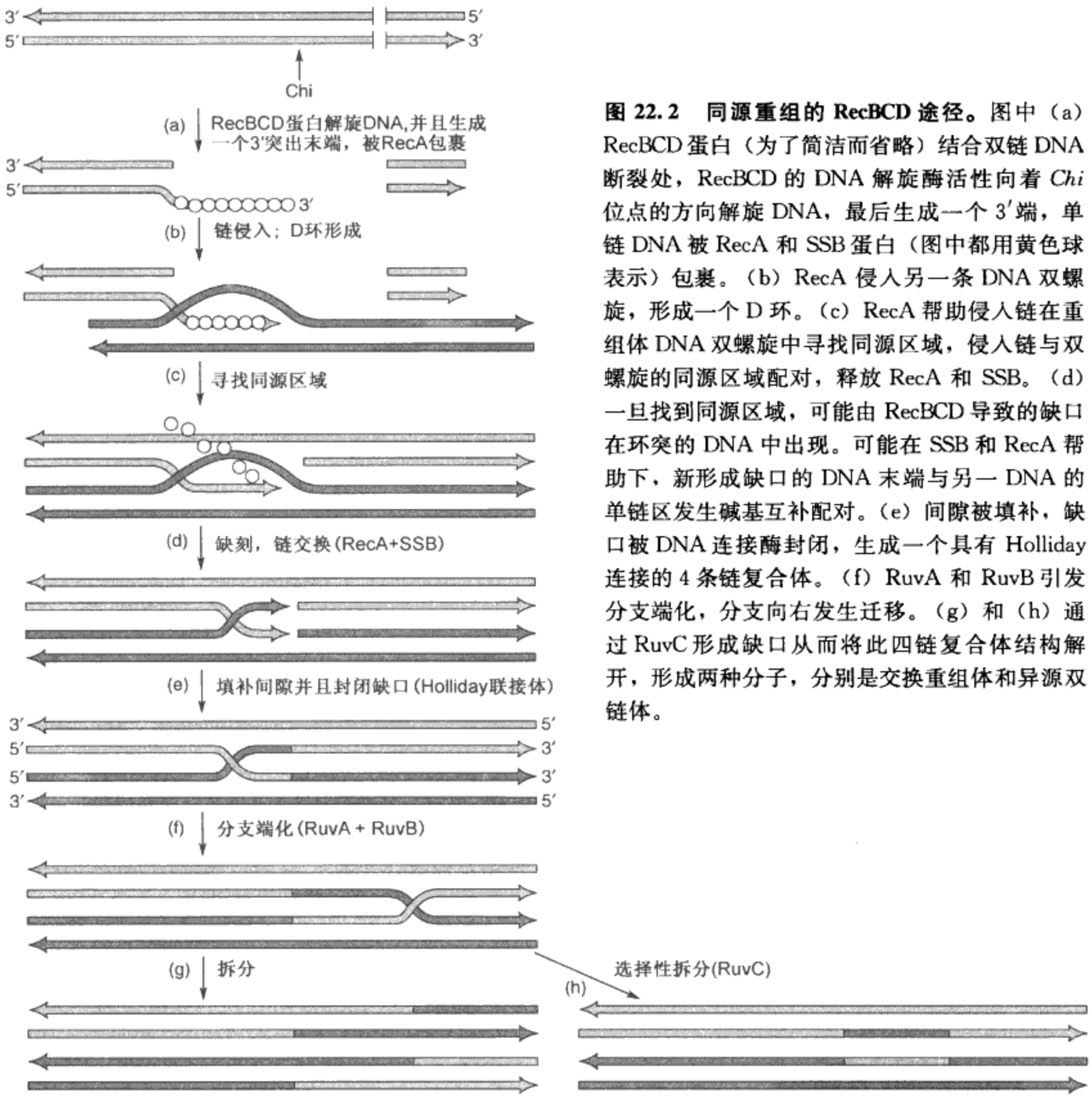


图 22.2　同源重组的 RecBCD 途径。图中（a）RecBCD 蛋白（为了简洁而省略）结合双链 DNA 断裂处，RecBCD 的 DNA 解旋酶活性向着 *Chi* 位点的方向解旋 DNA，最后生成一个 3′端，单链 DNA 被 RecA 和 SSB 蛋白（图中都用黄色球表示）包裹。（b）RecA 侵入另一条 DNA 双螺旋，形成一个 D 环。（c）RecA 帮助侵入链在重组体 DNA 双螺旋中寻找同源区域，侵入链与双螺旋的同源区域配对，释放 RecA 和 SSB。（d）一旦找到同源区域，可能由 RecBCD 导致的缺口在环突的 DNA 中出现。可能在 SSB 和 RecA 帮助下，新形成缺口的 DNA 末端与另一 DNA 的单链区发生碱基互补配对。（e）间隙被填补，缺口被 DNA 连接酶封闭，生成一个具有 Holliday 连接的 4 条链复合体。（f）RuvA 和 RuvB 引发分支端化，分支向右发生迁移。（g）和（h）通过 RuvC 形成缺口从而将此四链复合体结构解开，形成两种分子，分别是交换重组体和异源双链体。

我们还不知道这两种机制在体内是否发挥作用。但其中的任何一种都会形成具有自由 3′端的被 RecA 包裹的单链 DNA。这种结构对起始重组过程是必需的。

小结　由 *E. coli* RecBCD 引发的 DNA 同源重组过程起始于一个被 RecA 包裹的带有游离 3′端的单链 DNA，而这个单链 DNA 是来自另一个双链 DNA 受到切割后所释放出来的。这条侵入链形成一个 D 环。接下来 D 环的降解形成一个分支中间体。这个中间体中的分支移端产生一个两条链在同源染色体之间发生交换的 Holiday 联接体。最后，Holliday 联接体可以从两切口处拆分。产生两条带有异源补丁的非交换重组 DNA 或侧翼 DNA 区域发生交换的重组 DNA。

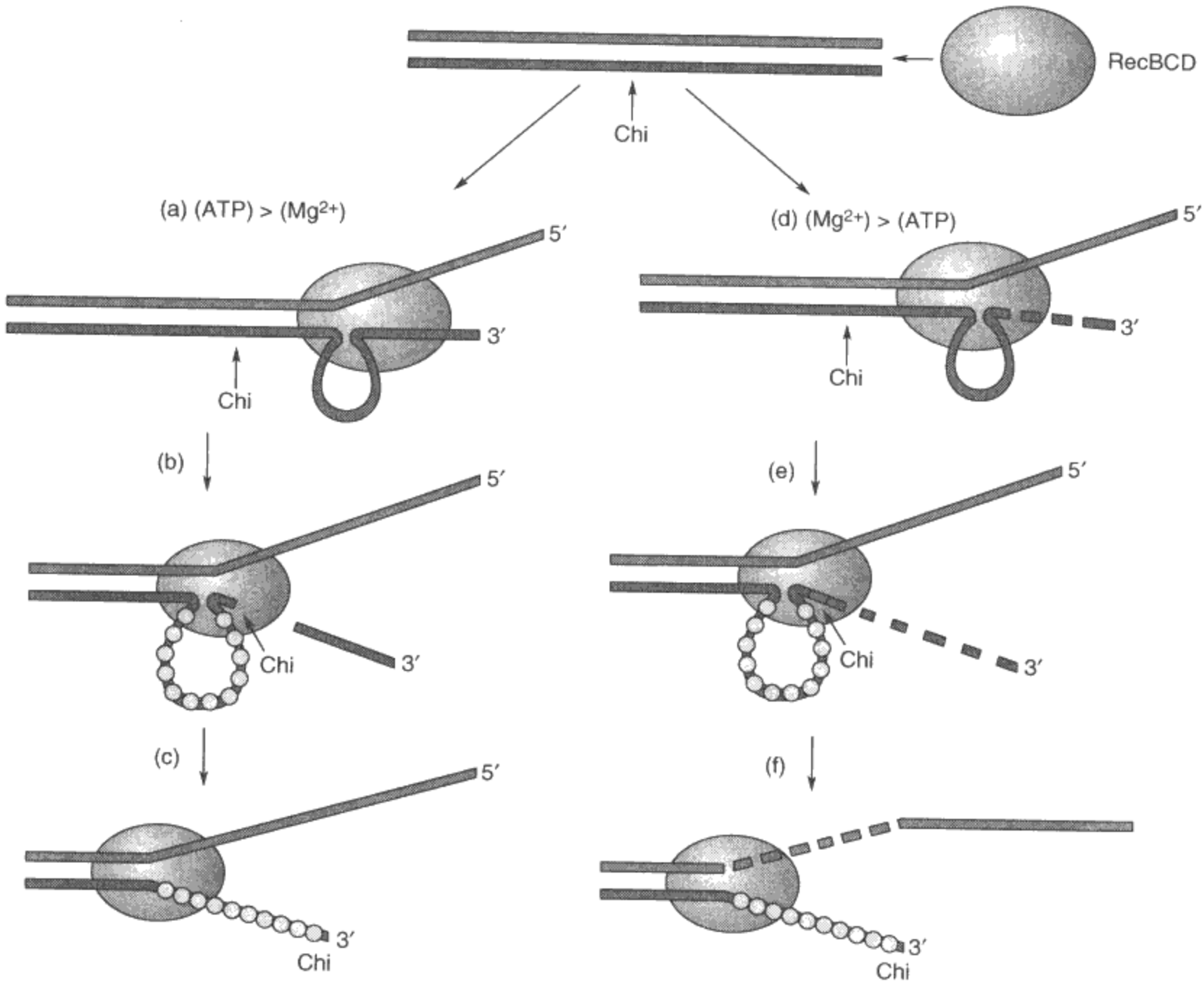


图 22.3　由 RecBCD 生成 3′端单链 DNA 的两种模型。左图描述了当体外 ATP 浓度大于 Mg^{2+} 浓度时，大多数的 Mg^{2+} 被 ATP 螯合。(a) RecBCD 结合到 DNA 平头末端并利用解旋酶活性开始解链 DNA，产生环状的 3′端链。(b) 当 RecBCD 接近 χ 位点时，在该位点的下游切割形成 3′端 DNA 链，将 RecA 蛋白负载到链的环形突出部分。(c) RecBCD 继续解链 DNA，形成 3′端覆盖着 RecA 蛋白的单链 DNA。右图描述了当体外 ATP 浓度小于 Mg^{2+} 浓度时，许多 Mg^{2+} 未被螯合。(d) RecBCD 结合到 DNA 平头末端并利用解旋酶活性开始解链 DNA，产生环状的 3′端单链区，在环后降解 DNA。(e) 当 RecBCD 接近 χ 位点时将 RecA 蛋白负载到链的环形突出部分。同时，RecBCD 停止降解 3′端链，而开始切割 5′端链，并激活 5′→3′外切核酸酶活性。(f) 在 RecBCD 继续解开 DNA 时，其利用 5′→3′外切核酸酶活性降解 5′端链，最终剩下一个 3′端覆盖着 RecA 的单链 DNA。(*Source*：Amundsen, S. K. and G. R. Smith, Interchangeable parts of the *Escherichia coli* recombination machinery. *Cell* 112 [2003], 743, p. 2. Reprinted with permission.)

22.2　RecBCD 途径的实验证据

我们已经简要地回顾了 RecBCD 途径，下面介绍一些支持这种重要的原核生物重组机制的实验证据。

RecA

在第 8 章讨论 λ 噬菌体的转导作用时已经介绍了 RecA。该蛋白质具有多种功能，但由于它最早是在研究重组时被发现的，因此而得名。早在 1965 年，Alvin Clark 和 Ann Dee Margulies 就分离到两种 *E. coli* 突变体，它们能够接受 F 质粒，但是不能通过重组永久性的将 F 质粒整合到宿主 DNA 中。这两种突变体对紫外线高度敏感，推测它们可能在采用重组机制修复紫外损伤方面有缺陷（见第 20 章）。对于这两个突变体的深入研究最终发现了两种 RecBCD 途径中的关键蛋白质：RecA 和 RecBCD。

对 *recA* 基因的克隆和超量表达，使对 RecA 蛋白的研究变得方便可行。RecA 是一个 38kDa 的蛋白质，在体外可以促进多种链交换的反应。通过体外实验，Charles Radding 等明确了 RecA 参与链交换的 3 个阶段。

1. 联会前。RecA 和 SSB 一起包裹参与重组的单链 DNA。

2. 联会。参与链交换的单链 DNA 或者双链 DNA 的互补序列的配对。

3. 联会后或链交换阶段。单链 DNA 替换双链中的正链 DNA，形成新的双螺旋。形成中间体，即联合分子（joint molecule）使得两条 DNA 链相互缠绕，开始链交换。

联会前（presynapsis） 最好的证据是可视性地观察到 RecA 能和单链 DNA 结合。Radding 等构建了一个带单链末端的线性双链噬菌体 DNA，将 RecA 与该 DNA 共同培养，通过电子显微镜观察这种复合体并照相。图 22.4a 显示了 RecA 特异结合单链末端，形成蛋白质包裹的 DNA 纤维，中间的双链 DNA 没有被包裹。同时培养 RecA 与单链环状 M13 噬菌体 DNA。图 22.4b 显示延展的 DNA 环均匀地被 RecA 包裹。图 22.4a、图 22.4b 两图的放大倍数是一致的，比较图 22.4a 中裸露 DNA 与 RecA-DNA 复合体的厚度就可以清楚地表明图图 22.4b 中环形 DNA 上确实包裹有 RecA。

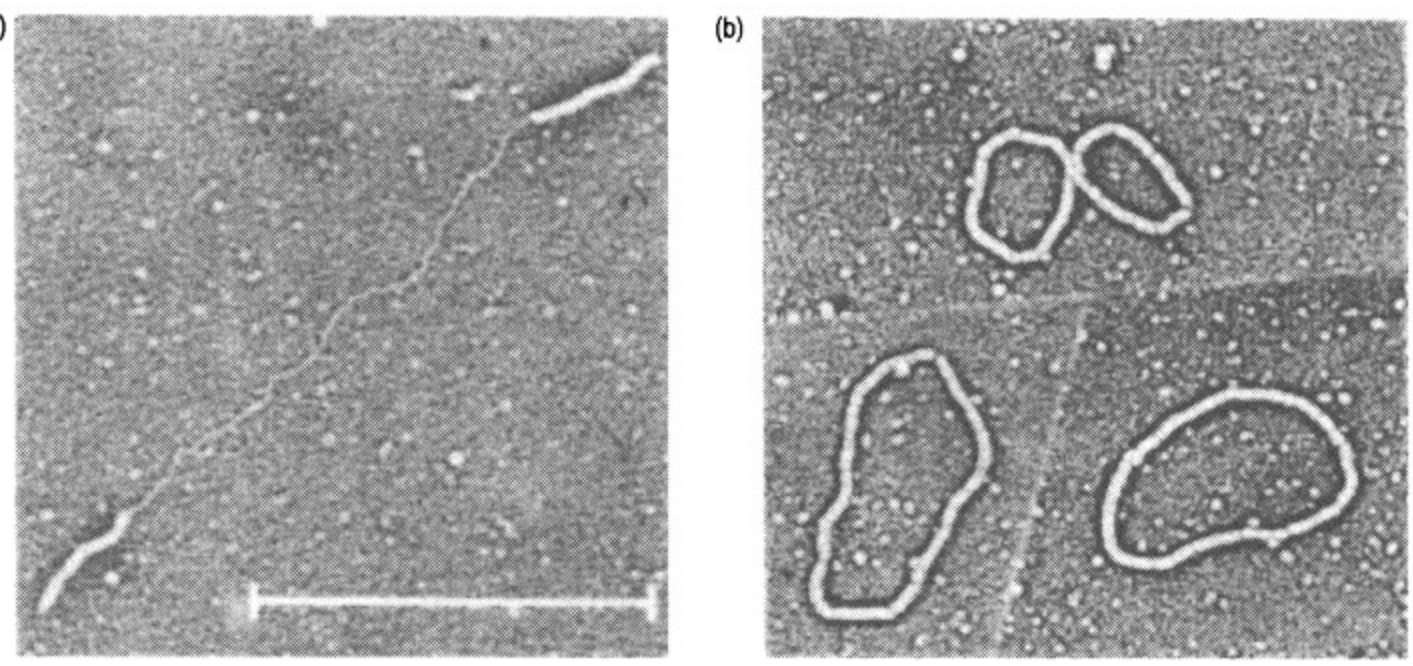


图 22.4 RecA 蛋白结合到单链 DNA 上。Radding 等制备了（a）具有单链末端的线型双链噬菌体 DNA 和（b）环状单链噬菌体 DNA。接着添加 RecA，经过一定时间，形成复合体后，在包埋的电子显微镜栅格上展开复合体，然后拍照。（a）图中的标尺为 500nm，（b）图相同。（*Source*: Radding, C. M., J. Flory, A. Wu, R. Kahn, C. DasGupta, D. Gonda, M. Bianchi and S. S. Tsang, Three phases in homologous pairing: Polymerization of *recA* protein on single-stranded DNA, synapsis, and polar strand exchange. *Cold Spring Harbor Symposia of Quantitative Biology* 47 (1982) f. 3 f& j, p. 823.）

有实验清楚地表明在联会前期形成包裹性 DNA 的过程中，SSB 发挥着重要的作用。Radding 等表明单链 M13 噬菌体 DNA 单独和 SSB 混合，与 M13DNA+SSB+RecA 一起混合所形成的 DNA 蛋白质复合体是显著不同的。图 22.4 表明 M13 噬菌体 DNA 与 SSB 和 RecA 形成的 DNA 蛋白质复合体比单独与 RecA 生成的复合体坚固。此外，Radding 等表明 SSB 能加速 DNA 的包裹。在 SSB 和 RecA 同时存在时，这些被包裹的 DNA 在 10min 内就形成。相反如果在没有 SSB 时，该种 DNA 纤维在 10min 后才会开始形成。

既然 RecA 可以单独包裹单链 DNA，SSB 的作用是什么呢？SSB 可能在解开 DNA 二级结构，如发夹结构等方面发挥作用。有多个实验证实了这一观点。Radding 等在高浓度的和低浓度的 $MgCl_2$ 的条件下，将 RecA 与单链共同培养时，测定链交换。发现低浓度的 $MgCl_2$ 降低了 DNA 二级结构的稳定性，而高浓度的 $MgCl_2$ 则会增加 DNA 二级结构的稳定性。该实验中，在高 $MgCl_2$ 浓度下 SSB 是必需的，而在低 $MgCl_2$ 浓度下 SSB 不是必需的。这与我们所设想的 SSB 在松弛 DNA 二级结构中是必需的结果是一致的。

在本节的后面部分，我们将看到在链交换中 ATP 的水解作用也是非常必要的。但 L. R. Lehman 等的结果表明在 SSB 存在的条件下，一种不具有水解作用的 ATP 类似物——ATPγS，在一定程度上也能维持链交换。基于 ATPγS 能导致 RecA 不可逆地与单链或者双链 DNA 紧密结合的事实，Radding 研究小组提出一个假设，认为 ATPγS 能够限制 DNA 二级结构的形成从而使链交换得以进行。如果假设是正确的，那么在 RecA 与单链 DNA 结合之前，SSB 先与 DNA 结合，便可以去除单链 DNA 中的二级结构，从而有利于链的交换。正如预期的一

样，SSB 优先于 RecA 与 DNA 结合，并确实加速了链的交换。这些都证明 SSB 通过打开二级结构而促使单链 DNA 参与重组过程。

小结 在重组的联会前阶段中，RecA 包被参与重组的单链 DNA。单链 DNA 结合蛋白（SSB）通过解除 DNA 二级结构，避免了因 RecA 蛋白与任一 DNA 二级结构结合对后续链交换所产生的抑制作用，从而加速了重组过程。

联会（synapsis）：互补序列的结合 我们接下来将会介绍 RecA 促进 DNA 双链中的一条链侵入到另一双链 DNA 中的交换过程。在这一过程中，侵入链与另一个双链 DNA 的其中一条链形成一个新的双链结构。这一步骤虽先于链交换、联会，但它发生了互补序列的简单配对而没有形成相互缠绕的双螺旋。由于在联会过程中形成的产物不稳定，所以相对于链的交换，联会过程是一个难以检测到的阶段。然而，Radding 等早在 1980 年就提供了很好的联会阶段的实验证据。

就像在研究联会前阶段的实验一样，在观察联会过程的实验中，电子显微镜是应用得最多的关键技术。Radding 等利用最熟悉的实验材料：单链环状噬菌体 DNA 和双链线性 DNA 进行联会研究。然而，在该实验中所应用的单链环状 DNA 是 G4 噬菌体 DNA，双链线性 DNA 是接近中间部位插入了一个 274bp G4 DNA 的 M13 噬菌体 DNA。由于该单链 G4 DNA 的靶位点序列有几千个碱基，并且缺口很少，所以发生链交换的可能性极小，相反，互补序列的简单联会却会发生，如图 22.5 所示。

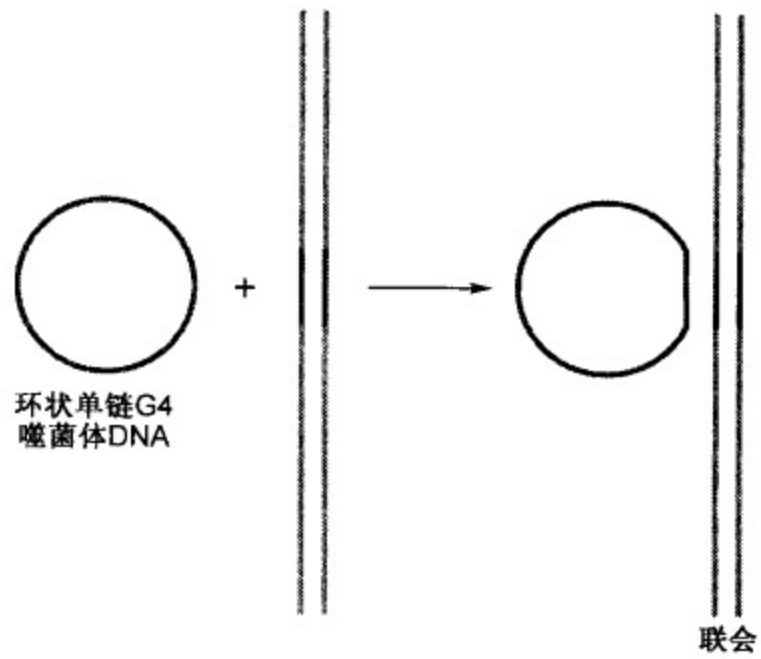


图 22.5 联会。图中显示了在环状单链 G4 噬菌体 DNA（红色）和线状双链 M13（G4）DNA（蓝色）[带有的一段 274bp 的 G4 噬菌体 DNA（红色）] 之间存在的联会现象。联会不包括任何线状和环状 DNA 的相互缠绕。

为了检测两个 DNA 分子间的联会，Radding 等分别在加入 RecA 和不加入 RecA 的情况下混合这两种 DNA 分子，并在电镜下观察。结果如图 22.6 所示，大部分的 DNA 分子已经发生联会。在大部分的联会分子中，配对区域的长度是适当的，配对区域的位置在线状 DNA 中是准确的。此外，在两分子间的非同源区域联会减少为原先的 1/40～1/20。另外在不存在 RecA 的条件下联会无法进行。

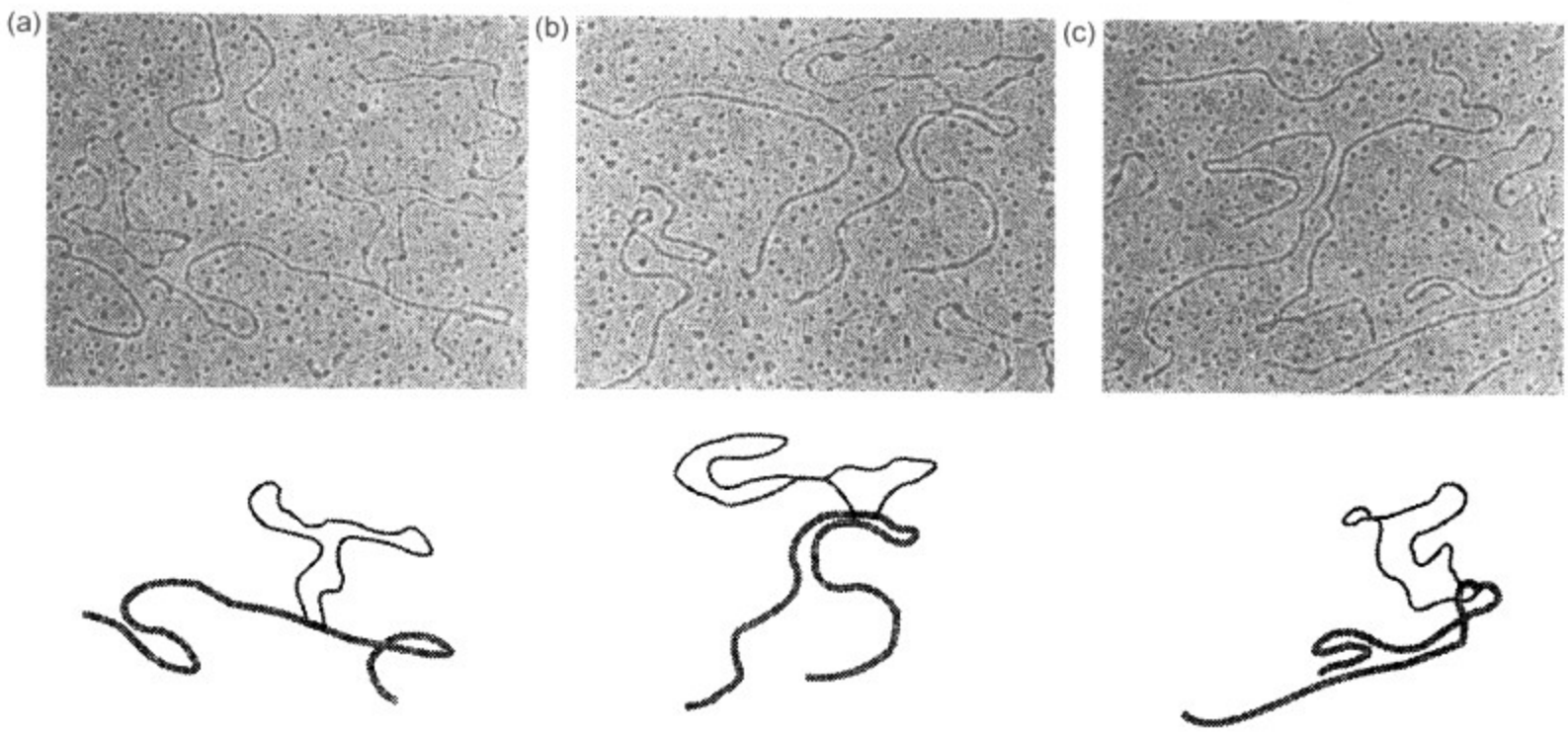


图 22.6 体外依赖于 RecA 的联会。Radding 等如图 22.5 混合环状单链 DNA 和线状双链 DNA 以及 RecA，电镜下检测联会产物。(a)、(b)、(c) 显示了 DNA 分子配对的 3 个实例。每个电镜图的下面是其示意图，蓝色代表线状双链 DNA，红色代表环状单链 DNA。粗红线指示每种情况下两 DNA 的联会区域。（*Source*: DasGupta C., T. Shibata, R. P. Cunningham and C. M. Radding, The topology of homologous pairing promoted by *recA* protein. *Cell* 22 (Nov 1980 Pt2) f. 9 d-f, p. 443. Reprinted by permission of Elsevier Science.）

尽管在这个实验中有缺口的双链 DNA 的比例很低，但是这些缺口在线状 DNA 中可以产生游离的末端，从而导致相互缠绕的具绞旋线双螺旋（plectonemic double helix）的结构形成。如果上述情况发生，那么两条 DNA 之间的连接即使是在接近 DNA 变性温度的条件下仍然是很稳定的。然而，如图 22.6 所示，在低于复性温度 20℃时加热 5min，配对结构就被破坏了。因此，这里所观察到的联会就不包括通过碱基配对形成的 Watson-Crick 双螺旋结构。相反，它可能包含一种由两个配对的 DNA 链并排而并非相互缠绕形成的具平行线的双螺旋（paranemic double helix）结构。这个结果进一步支持缺口在联会中不是必要的，在实验中，双螺旋 DNA 可以像线性双链 DNA 一样起作用。

发生联会需要多大程度的同源性呢？David Gonda 和 Charles Radding 估算 151bp 的同源区域与 274bp 的同源区域可以发生相同效率的联会。但是具有 30bp 的同源区域的两条 DNA 链，则只会出现基本水平的条 DNA 分子的配对。因此，有效联会的最小的同源度介于 30～151bp。

小结 单链 DNA 在双链 DNA 上发现其同源区域并与其配对时发生联会，在这种情况下两条 DNA 分子不发生缠绕。

联会后（postsynapsis）：链交换 在链交换的前两个阶段——联会前和联会阶段中，RecA 蛋白是必需的。而在链交换阶段我们也将会看到同样也需要 RecA。Lehman 等利用过滤结合实验分析 D 环的形成，测定双链和单链噬菌体 DNA 之间的链交换。实验方法如下：在存在 RecA 和不存在 RecA 的条件下，培养经 ^{3}H 标记的 P22 噬菌体双链 DNA 和未标记的 P22 的单链 DNA。利用限制分支移端的高盐浓度和低温条件以完全消化单链 DNA，消除 D 环。接着利用去垢剂（十二烷基肌氨酸钠或十二烷基磺酸钠）从 DNA 上洗去蛋白质。最后，通过硝酸纤维膜过滤。如果在双螺旋 DNA 中形成了 D 环，该单链 D 环将导致一种复合体结合在滤膜上，被标记的 DNA 将被保留下来。如果没有 D 环形成，那么未标记的单链 DNA 将结合到滤膜上，而标记的双螺旋 DNA 将会通过滤膜。去垢剂通过与 RecA 的结合而阻止 DNA 结合到滤膜上。Lehman 等同时也用双螺旋 M13 噬菌体 DNA 和线状 M13 噬菌体 DNA 进行上述实验。结果发现，在存在 RecA 的条件下大约 50%的 DNA 双螺旋形成 D 环。而不存在 RecA 时，小于 1%的 DNA 形成 D 环。同样，如果使用的是非同源的单链 DNA，也仅有 2%的 DNA 形成 D 环。

为了证实 D 环确实已经形成，该研究小组利用 S1 核酸酶消化这些复合体，去掉其中的单链 DNA，随后过滤产物。通过这种处理大大降低了标记 DNA 在滤膜上的滞留率，结果证实确实有 D 环的形成。实际上，在线性和超螺旋 DNA 中通过电镜可以直接观察到清晰可见的 D 环。这个实验表明超螺旋 DNA 对于链交换不是必需的。

表 22.1 总结了通过检测在硝酸纤维素滤膜中的滞留情况，揭示不同的核苷酸对于 D 环形成的影响。其中 ATP 对于 D 环的形成是必需的，它的作用不能被 GTP、UTP 或者 ATPγS 所替代。ATPγS 不能代替 ATP 的事实表明在 D 环形成中，ATP 的水解作用是必需的。事实上，ATP 的水解作用使得 RecA 与 DNA 分离，进而促成在链交换过程中必须发生的新的碱基配对。

表 22.1 D 环形成所需要的因子

双链 DNA	互作组分	D 环形成比率/%
P22 噬菌体	完整的	100
	-RecA	＜1
	-ATP	＜1
	-ATP+GTP	＜1
	-ATP+UTP	＜1
	-ATP+ATPγS	＜1
M13 噬菌体	完整的	100
	-RecA	1
	-ATP	1
	-ATP+GTP	2

据此，这个实验证明了 ATP 的水解作用对于 D 环的形成是非常重要的。RecA，一个难以置信的万能蛋白质，具有 ATP 酶活性，当它从 DNA 上分离时水解 ATP，促进 D 环的形成。

小结 RecA 和 ATP 共同启动单链与双链 DNA 间的链交换。为了使 RecA 从联会的 DNA 上分离，并促使一条单链和来自一双链 DNA 中的另一单链结合形成新的双链 DNA，ATP 是必需的。

RecBCD

我们在讨论 RecA 的过程中，已经涉及单链和双链 DNA 反应模型。这是因为 RecA 需要一单链 DNA 起始链交换。但是在自然情况下发生的重组常常涉及的是双链 DNA。RecA 怎样得到它所需要的单链 DNA 呢？答案是，它是由 RecBCD 蛋白提供的。这个过程涉及两个因子，分别是 DNA 上的 *Chi* 位点和 RecBCD 的解旋酶活性。下面列举有关这两个因子的实验证据。

Chi 位点是在细菌 λ 噬菌体的遗传研究中发现

的。λ *red gam* 噬菌体缺少 *Chi* 位点，但是它们的有效复制又依赖于 RecBCD 途径的重组。由于 RecBCD 途径依赖于 *Chi* 位点，因此这些突变体只能形成很小的噬菌斑。Franklin Stahl 等发现某些 λ *red gam* 噬菌体的突变体可以产生大的菌斑，结果表明这些突变体存在很活跃的 RecBCD 重组。Stahl 等又发现在突变点附近重组被加强了，将这些突变产生的 *Chi* 位点称为“交换热点激活子”。在突变位点附近出现促进重组的事实提示，这些突变并不像一般发生在编码区的突变那样改变了基因产物的结构，相反，这些突变能产生新的 *Chi* 位点从而促进了附近的重组。

我们已经知道 *Chi* 位点能刺激 RecBCD 途径，而不是 λ Red（同源重组）途径或者 λ Int（位点特异重组）途径，或 *E. coli* 的 RecE、*E. coli* 的 RecF（均为同源重组）途径。这一点明显表明 *Chi* 位点参与到 RecBCD 途径之中，因为该位点是在 RecBCD 途径中唯一未在其他途径中被发现的组分。事实上，由于 RecBCD 具有内切核酸酶活性，因此由它切割 *Chi* 位点附近的 DNA 形成裂口来起始重组已成为一个很值得关注的假设。

Gerald Smith 等发现了支持这个假设的实验证据。他们制备了一个 3′端被标记的双链 DNA 质粒 pBR322，该质粒的末端附近有一个 *Chi* 位点。如图 22.7a 所示，3′端的标记离 *Chi* 位点仅 80bp。接着加入纯化的 RecBCD 蛋白。在 DNA 分子经过热变性后，其产物经电泳就可能发现由于 *Chi* 位点切割而产生的 80nt 的片段（必须很快地进行反应以避免 RecBCD 的非特异性核酸酶活性降解 DNA）。图 22.7b 表明确实有 80nt 的产物出现，并且它的出现与否依赖于 RecBCD 蛋白的有无。DNA 的热变性对产生 80nt 单链 DNA 是必需的，这也表明 RecBCD 蛋白不仅切割 DNA，而且还会越过缺口打开 DNA 分子。Smith 研究组利用 Maxam-Gilbert 测序法对同样的标记底物进行化学切割后得到的序列与标记的 80 个核苷酸片段一起电泳从而确定了精确的切割位点。他们观察到相差 1 个核苷酸的两条带，这提示 RecBCD 蛋白在底物的 2 个位点（即下列序列中用箭头所标识的位点）实施了切割。

5′-GCTGGTGGGTT* G* CCT-3′

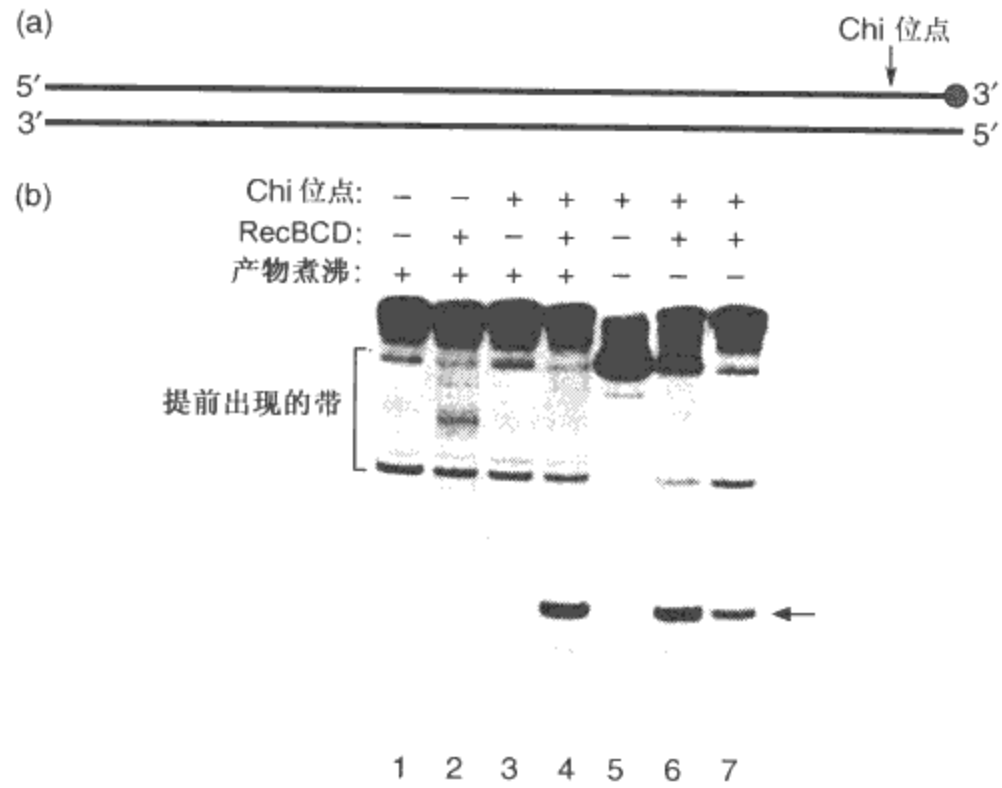


图 22.7 RecBCD 产生 DNA *Chi* 位点特异性的缺口。（a）底物缺口实验。Smith 等制备了一段 1.58kb *Eco*RI-*Dde*I 限制片段，这个片段具有距离 *Dde*I 末端大约 80bp 的 *Chi* 位点。其 3′端为 ^{32}P 标记的 *Dde*I 端（红色）。（b）缺口实验。Smith 等培养图（a）中的末端标记的 DNA 片段，在顶行中记做＋，而一个缺失 *Chi* 位点的相似片段，在顶行记做－。中间一行中，添加 RecBCD 或没有 RecBCD，分别记做＋或者－，反应 30s 后结束，产物进行电泳。产物如第三行所示，有些煮沸 3min（＋），有些没有（－）。右边的箭头指示着 *Chi* 位点形成缺口释放出来的 80nt 的标记片段。这个标记片段产物的出现依赖于 RecBCD 和一个 *Chi* 位点，而与是否煮沸产物无关。（*Source*：(*b*) Ponticelli, A. S., D. W. Schultz, A. F. Taylor and G. R. Smith, Chi-dependent DNA strand cleavage by *rec*BC enzyme. *Cell* 41(May 1985)f. 2, p. 146. Reprinted by permission of Elsevier Science.）

因此，对于上述底物，RecBCD 切割部位是距离其 *Chi* 位点 GCTGGTGG（画线部分）3′端的第 4 个或 5 个核苷酸处。对于其他的底物，切割部位可能是距离其 *Chi* 位点 3′端第 4 个、5 个或 6 个核苷酸处。因此，具体的切割位点因底物的不同而异。

这些结果支持 RecBCD 切割 *Chi* 位点附近的 DNA，也表明 RecBCD 能从缺口处解旋 DNA。Stephen Kowalczykowski 等的实验进一步支持 RecBCD 可以解旋 DNA。图 22.8 展示了他们其中一个实验的结果，表明：①单独的 RecA，或者 RecA 加 SSB，不能引起两同源双链 DNA 间的配对。②然而，有 RecBCD、RecA 和 SSB 时，只要两条 DNA 分子是同源的，依赖于 DNA 解旋的链交换就迅速发生。③如果其中一个 DNA 是经过热变性的，则 RecBCD 就不再是那么重要了。这最后一个发现暗示了 RecBCD 的一个功能是松弛 DNA 提供游离 DNA 末端；RecA 和 SSB 能包裹此单链 DNA 并且利用其 DNA 末端发动链的入侵。

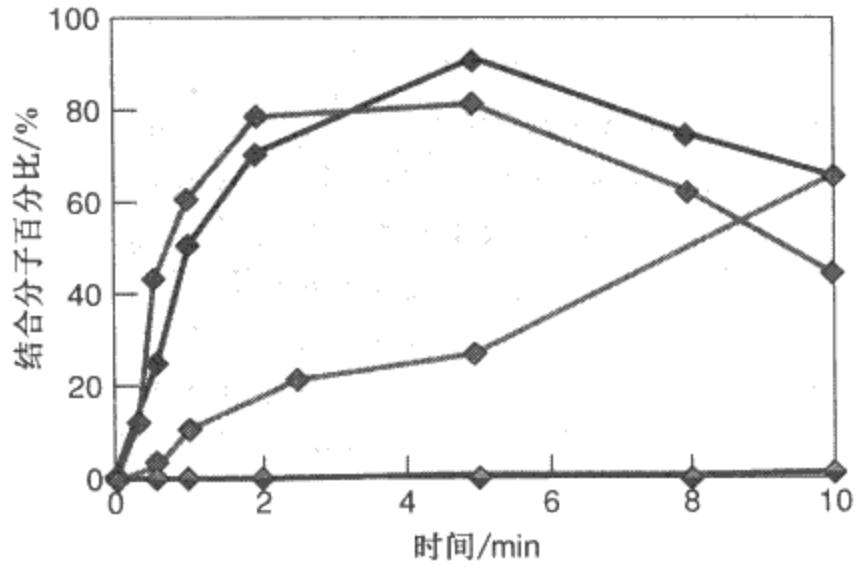


图 22.8　在两条双链 DNA 之间依赖 RecBCD 的链交换。Kowalczykowski 等将双螺旋 DNA 与 RecA、RecBCD、SSB（红色所示）共培养，并通过过滤结合或凝胶电泳分析结合分子（链交换）（结合分子比没有重新组合的 DNA 有更低的电泳迁移率）。还在缺乏 RecA 或缺乏 RecBCD（橙色或者紫色标记）的条件下分析了结合分子。蓝线表示其中一个 DNA 分子被热变性后，去除 RecBCD 的结果，其中一条 DNA 是热变性的。绿线表示除了 RecA 之外的所有组分预培养后，再添加 RecA 来启动反应的结果。其他所有的反应均是最后添加 RecBCD（*Source*：Adapted from Roman, L. J., D. A. Dixon and S. C. Kowalczykowski, "RecBCD-dependent joint molecule formation promoted by the *Escherichia coli* RecA and SSB proteins," *Proceedings of the National Academy of Sciences USA* 88: 3367-71, April 1991.）

Stuart Linn 等利用电镜检测松散的 T7 噬菌体 DNA 产物，为证明 RecBCD 具有 DNA 解旋酶活性提供了直接的证据。在实验中，当同时加入 SSB 和 RecBCD 时，他们观察到在毗邻的两个单链的双链 DNA 处具有分叉的结构，这暗示着 RecBCD 开始在双螺旋的末端解旋，SSB 包裹了两条已经形成的单链 DNA。而且正如预期的一样，随着时间的推移，分叉结构也随之拓展 。

小结　RecBCD 的 DNA 内切核酸酶活性能在 *Chi* 位点附近切割 DNA 形成缺口，RecBCD 的 ATP 酶驱动的 DNA 解旋酶活性，能够从缺口部位解开双链 DNA，释放游离的单链 DNA 末端，随后 RecA 蛋白和 SSB 与之结合，启动单链 DNA 的侵入和链的交换。

RuvA 和 RuvB

RuvA 和 RuvB 结合形成一个 DNA 解旋酶，催化 Holliday 联接体发生分支端化。我们已经知道，Holliday 联接体可以在体外生成。事实上，它们是验证 RecA 在链交换过程中的作用时的副产物。最初人们认为，RuvA 和 RuvB 在 Holliday 联接体中是和 RecA 相互作用的。后来，Stephen West 等设计了一种方法可对 4 种合成的寡核苷酸链进行分析，如图 22.9 所示，这些寡核苷酸的序列能通过碱基配对形成 Holliday 联接体。

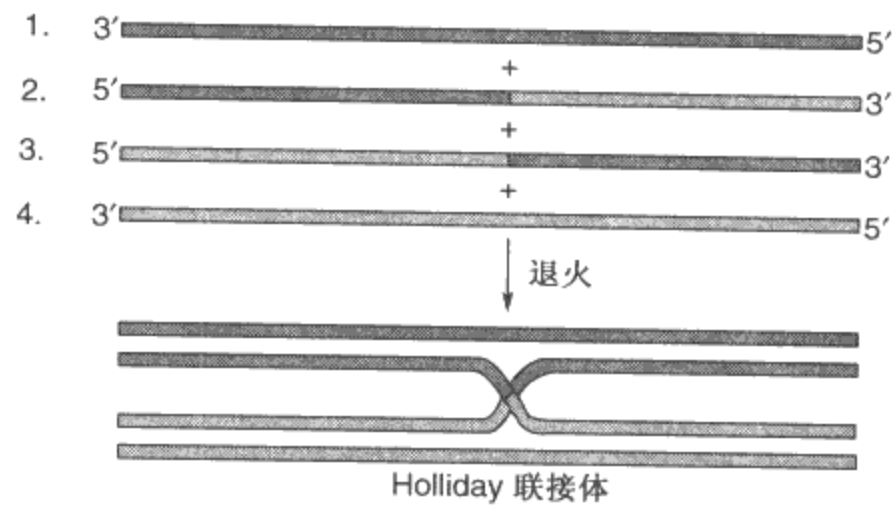
Carol Parsons 和 West 末端标记这个合成的 Holliday 联接体，并且利用凝胶阻滞实验来测定 RuvA 和 RuvB 与 Holliday 联接体的结合。由于交叉端化需要 ATP 提供能量，他们就使用了无水解能力的 ATP 类似物 ATPγS。理论上，RuvA 和 RuvB 将在 DNA 上结合，阻止分支端化，造成 Holliday 联接体拆分。该实验成功地证实了 RuvA 和 Holliday 联接体的存在，但是没有观察到与 RuvB 之间的移位，这表明已经形成了一个 RuvA-RuvB-Holliday 联接体复合体。表明这个三重复合体在这些实验条件下很不稳定。为了稳定假定的复合体，在试验中加入可以与复合体中的蛋白质交联的戊二醛，来阻止复合体在凝胶电泳中分离。

图 22.10 展示了 RuvA 和 RuvB 间的结合情况。在 RuvA 低浓度的条件下（泳道 b），只有很少的 RuvA结合到 Holliday 联接体 DNA 上。在 RuvA 高浓度的条件下，有大量的 RuvA/DNA 复合体产生。此外，仅 RuvB 本身，即使是在高浓度下，也不能与 Holliday 联接体 DNA 结合（泳道 e）。但是同时有两个蛋白质时，结合就可以发生，即使是在仅单独有 RuvA 又不能很好发生结合的低浓度下也能与 DNA 发生结合（泳道 f 和 g）。在缺乏 ATPγS 的情况下，单独 RuvA 可以结合 Holliday 联接体。但不能形成三重复合体 RuvA-RuvB-Holliday 联接体（泳道 h）。最后，RuvA、RuvB 或者两者都不能结合到与 Holliday 联接体相同长度的其他 DNA 双螺旋上（泳道 j～l）。所以这些蛋白质与 Holliday 的连接是特异性结合的。

在高浓度情况下，RuvB 自身能够驱动分支端化。这说明它具有 DNA 解旋酶活性和伴随的 ATP 酶活性。那么，RuvA 的作用是什么呢？RuvA 能结合到 Holliday 联接体的中心部位，促进 Holliday 联接体与 RuvB 的结合。因此，在很低浓度的 RuvB 浓度下，交叉移端也可以发生。此外，RuvA 还能促使 Holliday 联接体形成二维的正方形构象，该构象可使分支端化快速进行。

RuvA 和 Holliday 联接体间结合的特点是什么？David rice 等对 RuvA 四聚体进行 X 射线晶体学分析，结果表明它具有平面四方结构，从而支持了 RuvA-Holliday 联接体复合体是平面四方构象的假说。从图 22.11a 中可以看出，每个 RuvA 单聚体都

图 22.9　合成 Holliday 联接体的形成。寡核苷酸 1～4 混合变性，互补碱基互补配对。寡聚核苷酸 2 的 5′端（红色）与寡聚核苷酸 1（红色）的 3′端互补。这样两个分子的一半可以碱基互补配对，但寡聚核苷酸 2（红色）的 3′端与寡聚核苷酸 4 的 5′端互补（蓝色），产生碱基配对。寡聚核苷酸 3 的两末端与寡聚核苷酸 1 和 4 的另外一端互补，导致寡聚核苷酸 3 在与寡聚核苷酸 2 互补碱基配对时形成交叉。最后形成合成的 Holliday 联接体。



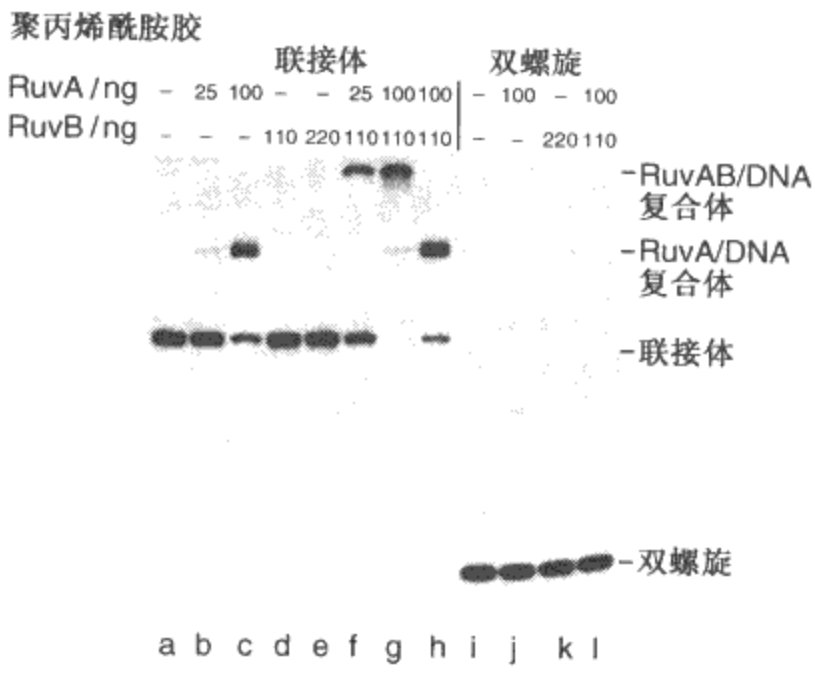


图 22.10　检测 RuvA-RuvB-Holliday 连接复合体。Parsons 和 West 将末端标记的合成的 Holliday 联接体，与不同数量的 RuvA 和 RuvB 混合，如上图所示，除了 h 泳道外，所有混合物中都包含 ATPγS（可提供能量，但无水解能力）。接着 Parsons 和 West 利用戊二醛使相同的复合体中的蛋白交联，防止这些蛋白质从复合体中解离。最后，将复合体进行聚丙烯酰胺凝胶电泳和放射自显影来检测标记的复合体。（*Source*：Parsons, C. A. and S. C. West, Formation of a RuvAB-Holliday junction complex *in vitro*. *Journal of Molecular Biology* 232（1993）f. 2, p. 400, by permission of Academic Press.）

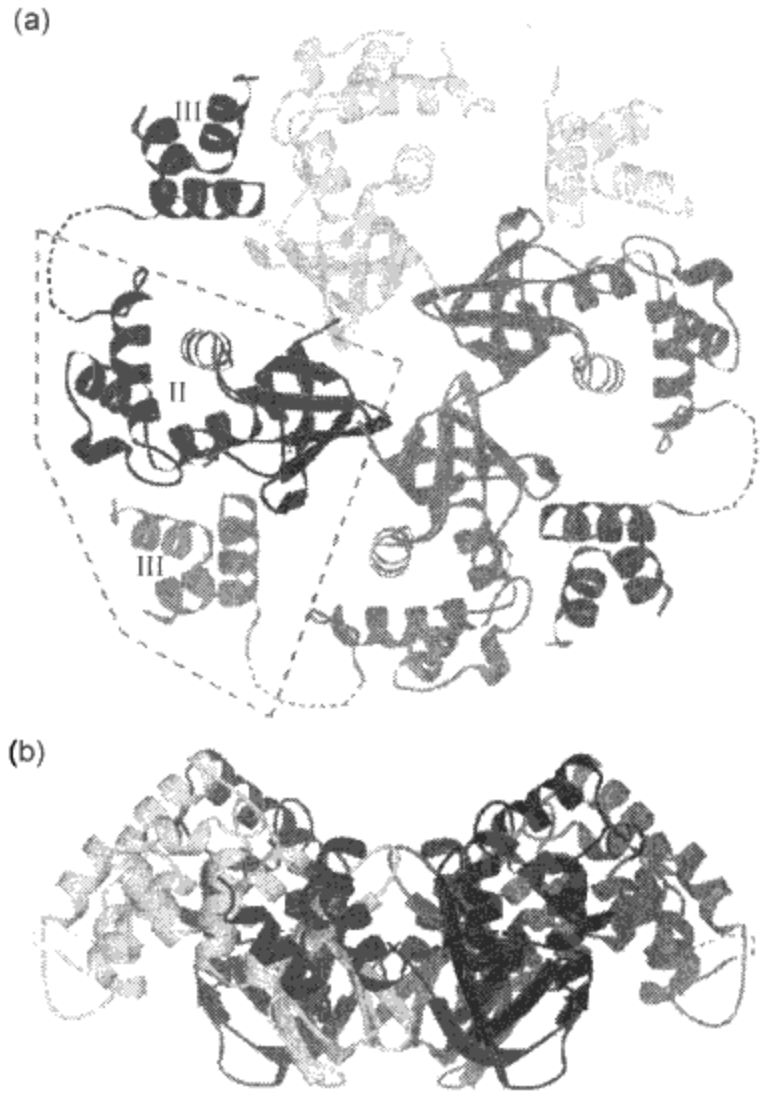


图 22.11　X 射线晶体衍射揭示的 RuvA 四聚体的结构。（a）顶面观，4 个单体分别由不同颜色显示。并且在这种二维立方结构的 4 个角中的 1 个，用 1 个红色虚线框标记。蓝色单体的 3 个结构域、绿色单体的第三个结构域（“L”形的“脚”）均以数字标出。（b）以相同的结构相同的颜色标记四聚体的侧面观，可以明显看出其顶部的凹面和底部的凸面。（*Source*：Rafferty J. B., S. E. Sedelnikova, D. Hargreaves, P. J. Artymiuk, P. J. Baker, G. J. Sharples, A. A. Mahdi, R. G. Lloyd and D. W. Rice, Crystal structure of DNA recombination protein RuvA and a model for its binding to the Holliday junction. *Science* 274（18 Oct 1996）f. 2 d-e, p. 417. Copyright © AAAS.）

是“L”形状，包含一个“腿”和“脚”及将其连接起来的不定形状的可变环（以短的彩色线表示）。每一个“L”形单体的“脚”与另一个相邻单体的“腿”缠在一起形成一个凸角，在图 22.11 中，4 个凸角的一个被红线框起来了。这些凸角排列以四聚物对称方式排列，每一对凸角之间有一个小沟结构。白色部分存在于两沟之间，图 22.11 b 显示了四聚体的侧面，揭示了顶部凹表面和底部的凸表面。

分子模型表明 RuvA 四聚体能在一致的平面四方构象里与 Holliday 联接体自然配对（图 22.12 所示）。DNA 和凹表面蛋白质之间的结合是很紧密的。Holliday 联接体的 4 个分支分布于蛋白质的表面的 4 个小沟中。4 个β转角分别位于 4 个单体上，构成了一个中空的发夹结构，从 Holliday 中间体的中间耸出。这个平面四方构象允许分支端化快速地进行。这个构象的任何偏离将减慢分支端化，这也强调了 RuvA 的平面四方构象的重要性。这种平面四方结构 Holliday 联接体与先前知道的相似的支链 Holliday 联接体之间的关系是什么呢？仅仅是同一结构的两种不同形式？下面我们就来了解。

Stephen West 和 Edward Egelman 与 Xiong Yu 一起对 RuvAB-Holliday 联接体复合体进行了进一步的电镜分析。他们将复合体的 100 个显微镜图扫描，结合产生一个标准图像。图 22.13a 展示了基于该标准图像的叠加有彩色 DNA 链的一个模型。结果与所期望的相似，RuvA 四聚体位于联接体的中间，两侧是两个 RuvB 的六聚体环。图 22.13b 显示将复合体的两臂弯曲的图像（为展示清楚而不显示 RuvA 四聚体）。图 22.13c 显示底部绕平面纸旋转 180° DNA 双螺旋的图像（RuvA 同样被移掉便于观察）。RuvB 环绕在这个相似的 Holliday 联接体上，通过箭头方向的移动来促进分支端化。尽管很难观察到，但 RuvB 也可以存在如图 22.13a 的形式。

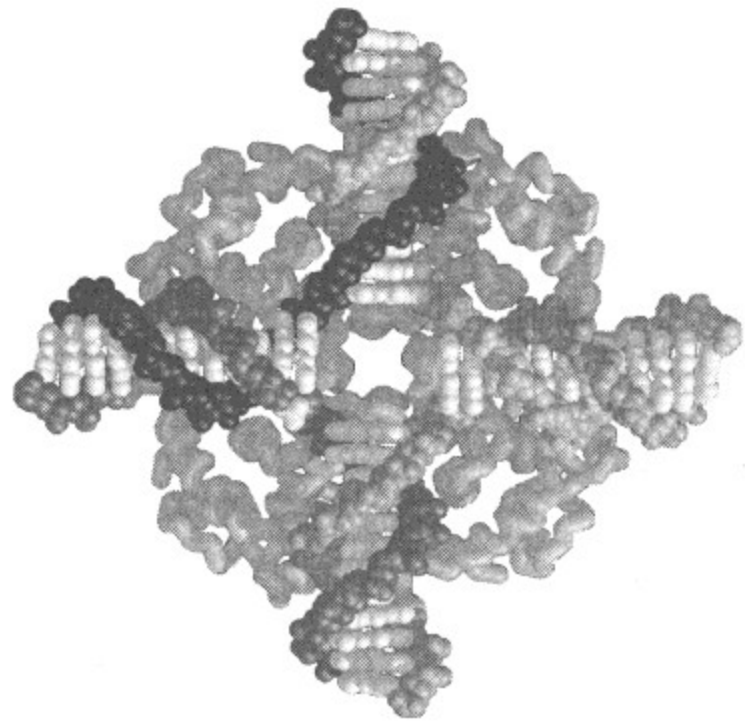


图 22.12　RuvA 和 Holliday 联接体之间的互作模型。RuvA 单体以绿色显示多肽链 α-碳骨架的管状结构，Holliday 联接体中的 DNA 由包含深红色和浅红色以及蓝色骨架和银色配对碱基的空间模型表示。黄色球显示能被 RuvC 切割，进而解开 Holliday 连接的两个配对位点中的一个的磷酸基团。（*Source*: Rafferty, J. B., S. E. Sedelnikova, D. Hargreaves, P. J. Artymiuk, P. J. Baker, G. J. Sharples, A. A. Mahdi, R. G. Lloyd and D. W. Rice, Crystal structure of DNA recombination protein RuvA and a model for its binding to the Holliday junction. *Science* 274（18 Oct 1996）f. 3d, p. 418. Copyright © AAAS.）

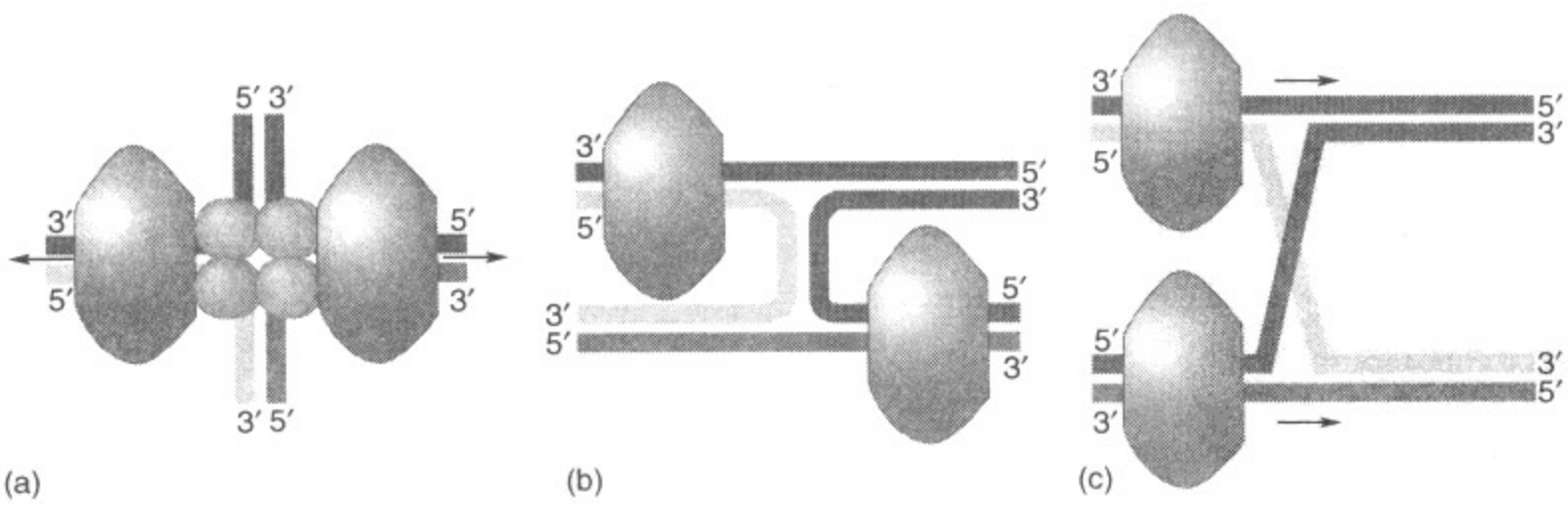


图 22.13　基于电子显微镜图像的 RuvAB-Holliday 联接体复合体模型。（a）复合体与 DNA 链之间正交垂直。DNA 链穿过复合体在箭头所指的方向移动。（b）图（a）中的蓝色、黄色和红色、绿色支链，向内旋转 90°。RuvA 四聚体去掉以便看到联接体的中心。（c）图（b）中的蓝色、绿色和蓝色、黄色链再向外旋转 180°。产生具有 RuvB 六聚体环的 Holliday 联接体，通过 RuvB 六聚体沿着箭头方向移动催化分支端化。（*Source*: Adapted from Yu, X., S. C. West and E. H. Egelman, Structure and subunit composition of the RuvAB-Holliday junction complex. *Journal of Molecular Biology* 266: 217-222, 1997.）

小结 由 RuvA 和 RuvB 构成的 DNA 解旋酶，驱动 Holliday 中间体的分支端化。其中 RuvA 四聚体具有平面的对称四方形结构，能够识别并结合于 Holliday 中间体。使 Holliday 中间体也采用了这种平面四方构象，从而启动在 Holliday 中间体直径方向上的两个对立的分支上结合 RuvB 的六聚体环。随后，RuvB 利用其 ATP 酶的活性驱动 DNA 螺旋解旋并重新形成新的螺旋。

RuvC

分支端化以后，Holliday 联接体如图 22.14 所示。正如我们在图 22.2 中所了解的有两种途径可以对联接体进行拆分。图 22.14 用另一种形式展现了这两种可能。在原始交换位置（位点 1，位点 2）切割两条链，将 Holliday 联接体拆分成两个带有异源双链 DNA 补丁的非交换的重组 DNA。另外，在位点 3、位点 4 发生切割，使另外两条链拆分成两条发生剪接的（交换的）重组 DNA。

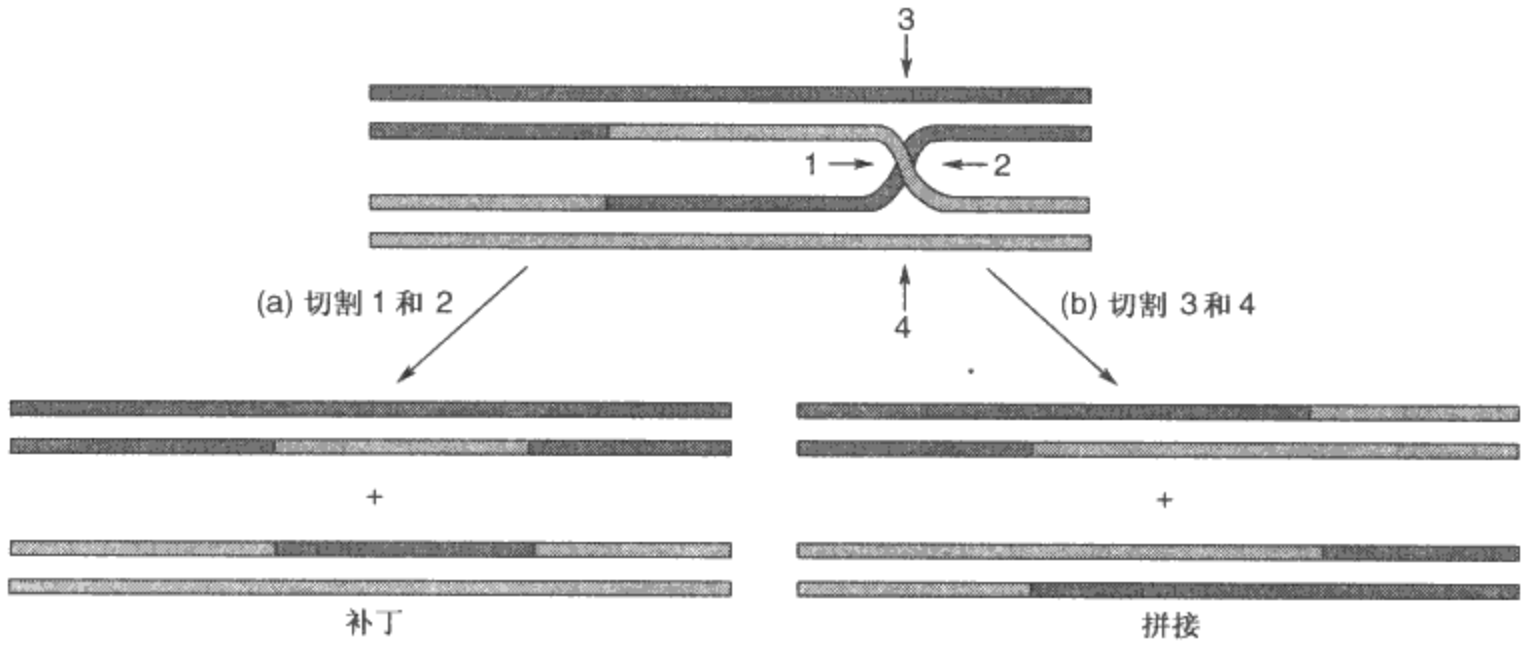


图 22.14 Holliday 联接体的拆分。图上方所示的 Holliday 联接体可依照图中带有编号的箭头，按两种不同的方式进行拆分。(a) 切割 1 和 2 产生带有“补丁”的异源双链 DNA。其长度与拆分前交叉迁移所涉及的相关位置的长度相关。(b) 切割 3 和 4 产生“拼接”片段的重组交换分子。

是哪一种酶通过酶切产生缺口来拆分 Holliday 联接体？West 等在 1991 年证明是 RuvC 行使了这一功能。如图 22.15a 所示，他们构建了经 ^{32}P 标记的 Holliday 联接体，在其联接点含有一个短的 12bp 的同源区域（J），这个结构的其他部分均由非同源区域构成。随后用凝胶阻滞实验检测 RuvC 与 Holliday 联接体或线型双螺旋 DNA 结合能力。图 22.15b 结果显示：当加入过量的 RuvC 到 Holliday 联接体中时，West 等观察到越来越多的 DNA-蛋白质复合体，证实了 RuvC 与 Holliday 联接体的结合。相同的实验也显示了 RuvC 不能与由链 1 及其互补链形成的线状双链 DNA 结合。

RuvC 可特异性地结合 Holliday 联接体，但是它能拆分这个联接体吗？图 22.15c 显示的答案是肯定的。West 等分别加入逐渐增加浓度的 RuvC 到被标记的 Holliday 联接体和线状双螺旋 DNA 的反应体系中。得到了所预期的结果，即 RuvC 拆分 Holliday 联接体产生带有标记的片段与线状双螺旋 DNA 长度相同，这也证实了 RuvC 导致 Holliday 联接体的拆分。通过更加复杂的实验可以区分带有补丁的产物与发生剪接的产物，结果表明在体外条件下，得到剪接产物的拆分方式为主要类型。

Kosuke Morikawa 等基于 X 射线晶体学研究，揭示 RuvC 的三级结构是一个二聚体，有两个相距 30Å 的活性位点。图 22.16a 所示，这种结构保证了在两个位点切割平面四方 Holliday 联接体。图 22.16b 对 RuvC-Holliday 联接体复合体进行了更细致的描述。RuvC 是像这个模型所描述的单独行使作用的，还是在 Holliday 联接体上与 RuvB 或 RuvB 加 RuvA 结合后作用于 Holliday 联接体？所获得的证据强有力地证明，RuvC 通过与 RuvA 及 RuvB 结合一起作用于 Holliday 联接体复合体行使功能。West 等重新构建了一个系统，在体外完成了从重组的中间产物到重组最后阶段的过程，结果表明 RuvC 与 RuvA、RuvB 的单克隆抗体阻断了 Holliday 联接的拆分。对于这个结果，可能的解释是 RuvA、RuvB 和 RuvC 三者相互作用。如果推论是正确的，这些蛋白质可以自然地结合在一起，并且其中的一个应该与其他的发生交联。West 等准备了两种不同蛋白质的混合物，加入戊二醛以实现交联，采用电泳方法检测交联效果。结果与预期的一样，RuvA 和 RuvB 可以交联，RuvB 和 RuvC 也可以交联，但是 RuvA 和 RuvC 之间不能交

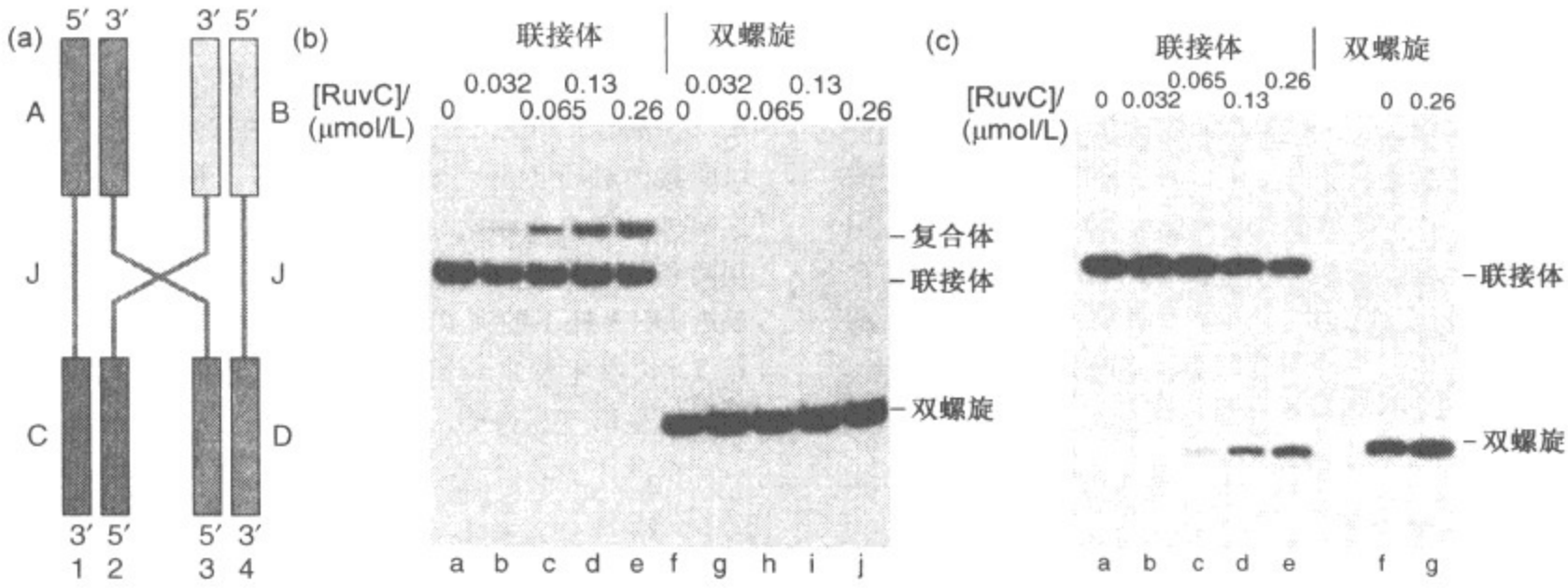


图 22.15　RuvC 拆分合成的 Holliday 联接体。（a）合成的 Holliday 联接体的结构。连接处具有一个 12bp 的短同源区（J 线段，红色），其他部分（A、B、C、D）是不同源的，以不同颜色表示。（b）RuvC 与合成的 Holliday 联接体的结合。West 等末端标记合成的 Holliday 联接体（用线性双螺旋 DNA 作对照），在非切割条件下（低温并缺乏 $MgCl_2$），与不等量的纯 RuvC 结合，电泳检测产物。结果表明 RuvC 结合到合成的 Holliday 联接体上，而不是普通的线性双螺旋 DNA 上。（c）RuvC 对 Holliday 联接体的拆分。West 等将标记的 Holliday 联接体（或线性双螺旋 DNA），在切割条件下（37℃，5mmol/L $MgCl_2$）与浓度不断升高的 RuvC 结合，电泳检测产物。结果表明 RuvC 将一些 Holliday 联接体拆分成线性双螺旋形式。（*Source*: Dunderdale, H. J., F. E. Benson, C A. Parsons, G. J. Sharples, R. G. Lloyd and S. C. West, Formation resolution of recombination intermediates by *E. coli rec*A and RuvC proteins. *Nature* 354 (19-26 Dec 1991) f. 5b-c. p. 509. Copyright © Macmillan Magazines Ltd.)

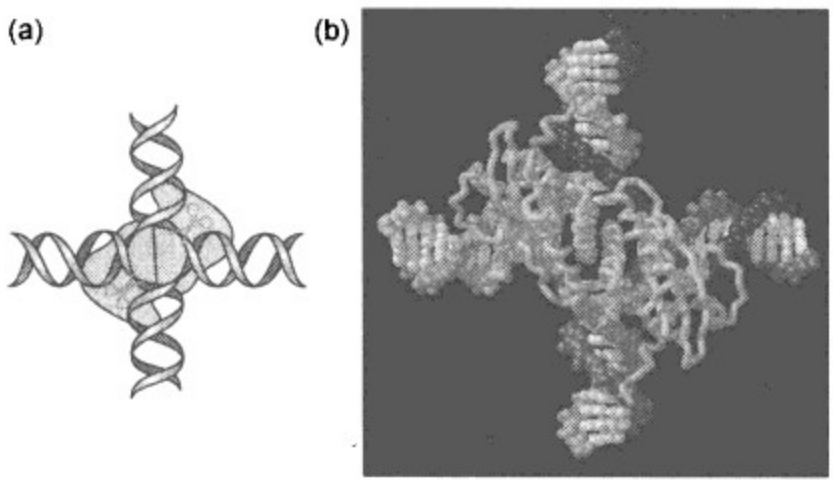


图 22.16　RuvC 和 Holliday 联接体的互作模型。（a）RuvC 二聚体（灰色）与二维四方形 Holliday 联接体结合的模型。“剪刀”（绿色）指示出两个 RuvC 单体上的活性位点（相距 30Å）。注意在拆分复合体中，RuvC 活性位点的位置与被切割的 DNA 的位置之间是如何契合的。（b）详细模型。灰色管代表 RuvC 二聚体碳骨架。Holliday 联接体中，蓝色和粉红色表示磷酸骨架，银灰色表示碱基对如图 22.12。（*Source*: Rafferty, J. B., S. E. Sedelnikova, D. Hargreaves, P. J. Artymiuk, P. J. Baker, G. J. Sharples, A. A. Mahdi, R. G. Lloyd and D. W. Rice, Crystal structure of DNA recombination protein RuvA and a model for its binding to the Holliday junction. *Science* 274 (18 Oct 1996) f. 3e, p. 418. Copyright © AAAS.)

联。可见，RuvB 能与 RuvA 和 RuvC 交联，表明 3 种蛋白质可以结合到一起与 Holliday 联接体结合。

RuvA、RuvB 和 RuvC 共同作用的假说与分支端化在拆分 Holliday 联接体中帮助 RuvC 发现首选切割位点的理论是一致的。这也和 X 射线晶体学研究显示的 RuvA 以四聚体或八聚体结合 Holliday 联接体相一致。West 假设包含 RuvA 八聚体的复合体对有效的分支端化是特异的（图 22.17a）。RuvC 可以取代 RuvA 四聚体中的一个，从而形成 RuvABC 连接复合体，即“拆分体”（图 22.17b），它对 Holliday 联接体的拆分是特异的。

ruvA、*ruvB* 和 *ruvC* 突变体均能形成相似的表型：由于缺失重组修复而对紫外光，离子辐射，抗生素丝裂霉素 C 高度敏感。但 RuvA 和 RuvB 具有促进分支端化的功能，而 RuvC 能催化 Holliday 联接体的拆分。为什么这 3 种蛋白的缺失却具有相同表型呢？答案是：可能拆分过程依赖于分支端化。RuvA 或 RuvB 突变后阻断分支端化，进而间接阻断 Holliday 联接体的拆分。

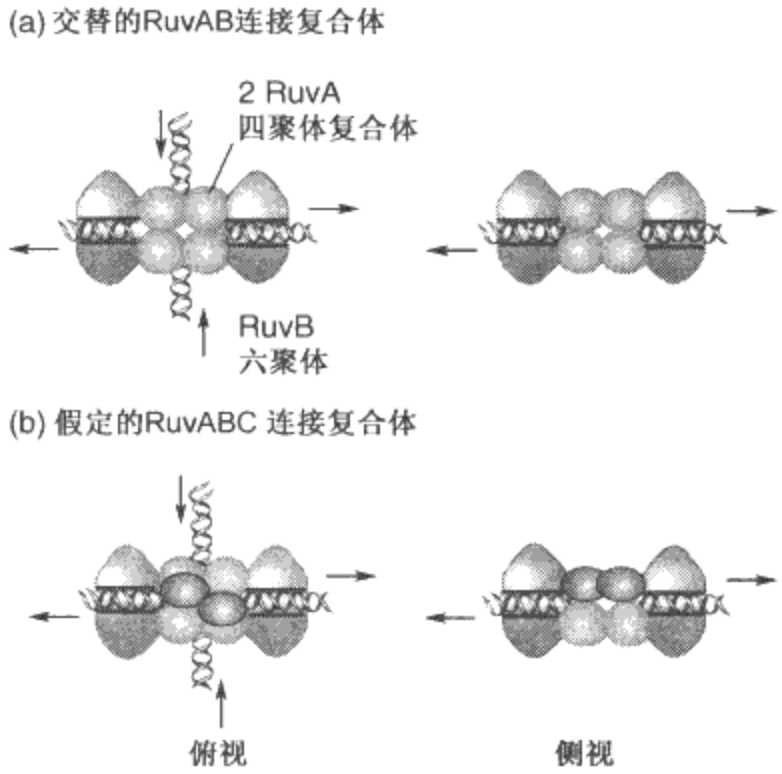


图 22.17　Ruv 蛋白-联接体复合体模型。(a) Pearl等发现了 RuvAB-联接体复合体。与其他研究者提出的 RuvA 四聚体复合体相比，该模型的 Holliday 联接体中包含一个 RuvA 八聚体。这个 RuvA 八聚体复合体可能是有效迁移时的一种形式。(b) West 的 RuvABC-联接体复合体模型。这个模型是 Holliday 联接体拆分过程中的一个复合体形式。(*Source*: Adapted from West, S. C., RuvA gets x-rayed on Holliday. *Cell* 94: 700, 1998.)

虽然 West 等并没有揭示确切的机制，但却阐明了拆分过程中 RuvC 酶作用的热区，而 RuvC 酶接近这些热区又需要相应的分支端化过程发生。为了确定 RuvC 切割位点的序列，West 又采用相同引物对相同的 DNA 进行 DNA 测序的引物延伸技术，对 RuvC 进行了分析 。结果，他们共检测到 19 个切割位点，找到一个明显的保守序列：5′-(A/T) TT↓ (G/C)-3′。据此推测，在活体内必须由 RuvA 和 RuvB 催化分支端化到这样一个保守区域。这个假设也意味着无论拆分成“异源补丁”或“拼接产物”，都依赖于两 DNA 链中 RuvC 识别的拆分保守序列的频率。总之，这应该是一个各占一半概率的拆分过程。

小结　Holliday 联接体的拆分是由 RuvC 酶催化的，这种蛋白质以二聚体的形式，从原始的交叉点处将两条链切开或形成两条非交叉的、带有“补丁”片段的异源重组双链 DNA，或切割另外两条链则会形成“拼接”的重组 DNA。这种蛋白质以二聚体形式将 DNA 夹住，优先在序列 5′-(A/T) TT↓ (G/C)-3′处产生剪切作用。“分支端化”到达优先的切割位点对于 Holliday 联接体的有效拆分是必需的。RuvA、RuvB 和 RuvC 酶在复合体中相互协调也直接参与“定位”与“切割”过程。

22.3　减数分裂重组

在本章的前面部分已经提到，大多数真核生物的减数分裂都伴随着重组。这个过程与细菌中的同源重组有很多相似的特征。在本节中我们将介绍酵母减数分裂重组的机制。

减数分裂重组的机制：综述

图 22.18 展现了已经研究得很透彻的酵母减数分裂重组的一个假设。该过程起始于染色体的损伤——双链断裂。接着，一个内切核酸酶识别这个断裂并且消化这两条链的 5′端，产生 3′单链凸出端。然后，这些单链末端侵入到 DNA 双螺旋形成与细菌同源重组中相似的 D 环结构。接着，DNA 修复合成填平了双螺旋的间隙，延伸 D 环来填充顶部双链。接着，在两个方向发生分支移端形成两个 Holliday 联接体。最后，Holliday 联接体被拆分产生一个包含两节异源链的非交换的重组体或产生一个交换了 DNA 侧翼区的交换重组体。

这个假设的大多数步骤已经得到了很好的实验证实，但也有一些实验与该假说的部分观点相抵触。一般而言，该模型预测杂合 DNA 会从双链 DNA 断裂处的两侧产生。然而，遗传学鉴定发现杂合 DNA 仅出现在缺口的一侧。只有少数情况能在断裂的双侧发现杂合 DNA。而且杂合 DNA 存在于相同染色单体上，而不是如模型预测的在两条染色单体上。因此，需要更多数据来解决这种偏差或修正这一假说。

双链 DNA 断裂

如何知道酵母中的重组是起始于双链 DNA 的断裂 (DSB) 呢？1989 年，Jack Szostak 等通过定位啤酒酵母 (*Saccharomyces cerevisia*) 中的 ARG4 基因的重组起始位点的基础研究回答了这个问题。他们在重组事件本身的过程中没有发现这一现象，但在减数分裂过程中的基因转换现象中发现了这一现象，而酵母的基因转换就依赖于减数分裂重组。由于基因转换和重组均起始于相同的位点，为此他们就通过研究基因转换机制来揭示重组的机制。我们将在下一节讨论基因转换的机制。

Szostak 等早期的工作表明 ARG4 位点的减数分裂的基因转换是有极性的。在靠近基因 5′端的转换是普遍的（占整个减数分裂中的 9%），而在 3′端的转换相对少见（占整个减数分裂的 0.4%）。这种特征表明重组的起始位点靠近基因的 5′端。因此，Szostak 等去掉这个区域，试图移除起始位点从而阻断基因

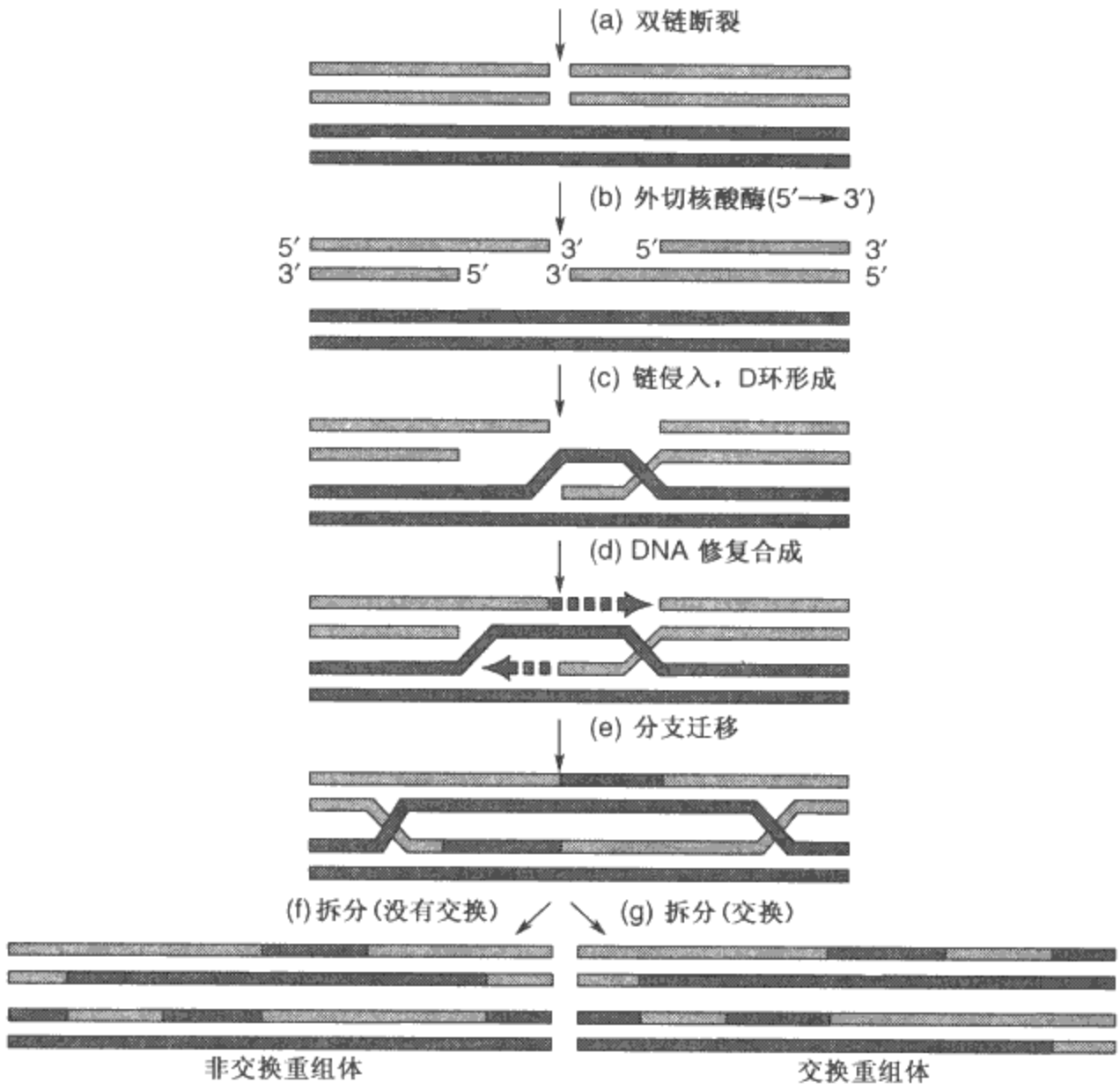


图 22.18　酵母减数分裂中的 DNA 重组模型。（a）在一条 DNA 双链（蓝色）中发生双链断裂，此链与另一条双链 DNA（红色）是配对的。（b）外切核酸酶在断裂处消化 DNA 的 5′端，产生新的断裂。（c）顶部双螺旋 DNA 游离出的 3′端单链侵入底部双螺旋 DNA 分子内，形成一个 D 环。（d）DNA 修复合成延伸的 3′游离末端，增大 D 环。（e）分支向两个方向移动，产生两个 Holliday 联接体。（f）Holliday 联接体内部链被切割而拆分，产生带有异源双链“补丁”的非交换重组体 DNA，但在 Holliday 联接体外的 DNA 臂上无交换。（g）切割左边的 Holliday 联接体的内部链和右边 Holliday 联接体的外部链，产生发生了交叉的重组 DNA 分子，该 DNA 具有右侧 Holliday 联接体右边的 DNA 交换区段。

转换。他们发现去除 ARG4 基因的−361 至+1 区显著降低了基因转换的频率，这也表明重组的起始位点存在于 ARG4 基因的启动子区。

这一信息能够帮助研究人员从酵母基因组的一些限制性区域寻找单链或双链 DNA 的断裂。因此，将克隆的一个长 15kb 包含 ARG4 基因的 DNA 片段，构建成质粒，并转化到一个酵母菌株中，该菌株一旦被转移到孢子生成培养基中即能同步完成减数分裂。在孢子形成过程的不同时间提取质粒 DNA，电泳，图 22.19 描述了电泳结果。在最初的时间，研究人员看到的多数是超螺旋单体、超螺旋的二聚体和松弛的环状单体，而低迁移率的带型可能是一些二聚体。在孢子生成诱导的整个时间里都出现了这些相同的带。孢子形成过程中有一条不明显的弱带，可能是线性单体，第一次在 3h 出现，4h 出现高峰，经过这些时间点后就逐渐消失。这些线状 DNA 应该是质粒的双链 DNA 发生断裂后产生的。DSB 发生的时间与在这些细胞中进行减数分裂重组的时间一致（2.5～5h），以及与重组产物出现的时间一致（4h）。这些结果都与减数分裂重组的假说中关于重组的第一步是 DSB 形成的观点相一致。

Szostak 研究组利用限制性作图证明 DSB 发生于质粒的 3 个不同位置，对于这一点，图 22.19a 中用箭头加以标识。这些断裂位点中的一个（位点 2）只存在 ARG4 基因的 5′调控区的一个 216bp 限制片段内。Szostak 等的前期工作表明在同一区域有 142bp 的缺失降低了 ARG4 基因的减数分裂期的基因转换。因此他们在这一研究中测试了相同的缺失对 DSB 的影响。结果他们发现这个缺失的确能在位点 2 消除 DSB，但不影响 1 和 3 位点的 DSB。因此，位点 2 形成 DSB 的能力与 ARG4 基因下游的减数分

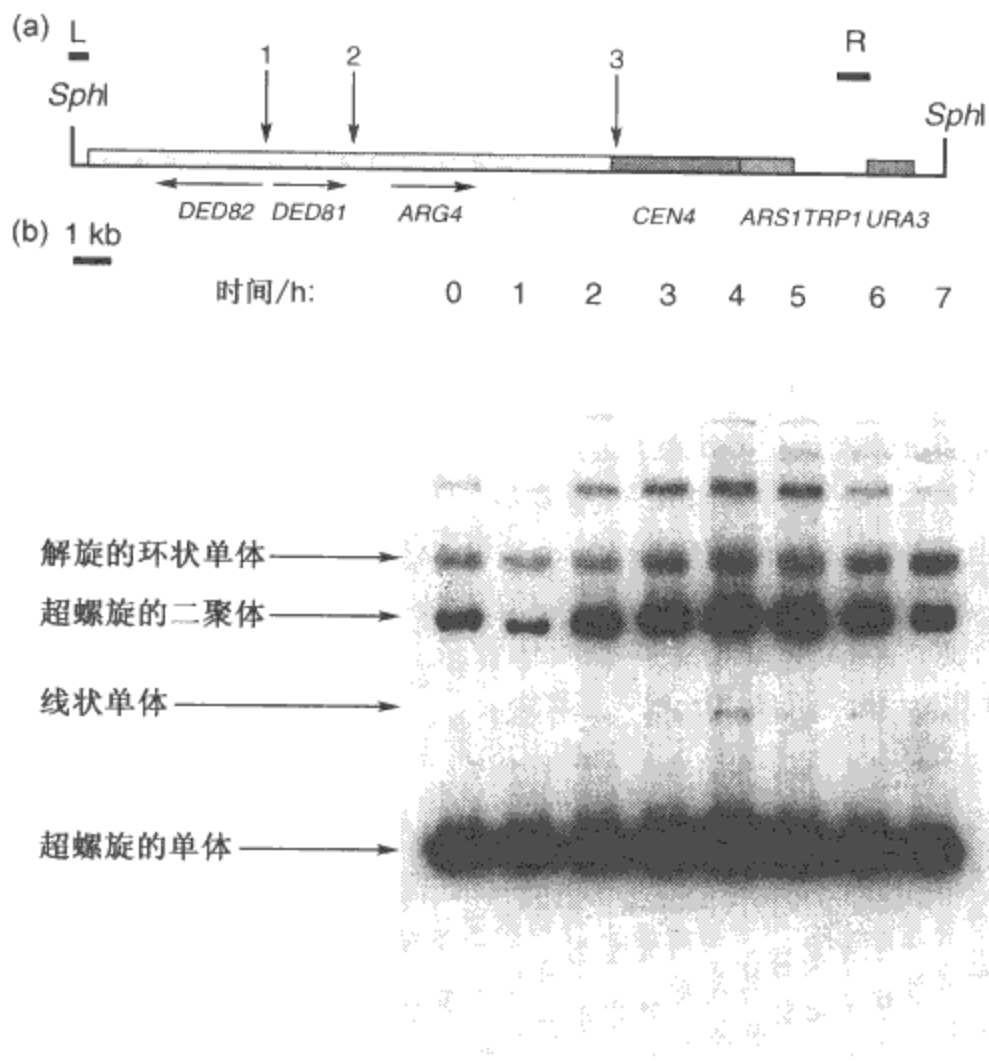


图 22.19　在具有重组起始位点的质粒上检测双链 DNA 断裂。（a）用来检测双链 DNA 断裂（DSB）的质粒图。黄色横条代表酵母基因组中包含重组起始位点的 15kb 的插入。其他颜色的横条分别代表载体内包括着丝粒（CEN4）在内的其他位点。基因的位置在横条下标出，L 和 R 是用来进行斑点杂交的探针的位置。箭头 1、2、3 所示的是质粒发生双链 DNA 断裂的 3 个不同位置。（b）电泳结果。Szostak 等将图（a）中描述的质粒转化到酵母中，诱导孢子形成，在诱导后的不同时间提取质粒 DNA 进行电泳，用^{32}P标记的质粒探针进行 Southern 杂交，检测 DNA。各种条带在左边注释。值得注意的是，诱导后 3～4h 出现一条新的相对较弱的线状单体 DNA，这种线状单体是双链 DNA 发生断裂产生的。（*Source*：Sun，H.，D. Treco，N. P. Schultes and J. W. Szostak，Double-strand breaks at an initiation site for meiotic gene conversion. *Nature* 338（2 Mar 1989）f. 1，p. 88. Copyright © Macmillan Magazines Ltd.）

裂基因转换的效率有关。

如果 DSB 的形成是减数分裂重组的起始条件，那么在酵母染色体中也会发生，而不会仅仅发生在质粒中。Szostak 研究组利用限制性图谱在缺乏质粒的细胞中寻找相同的 DSB。他们在酵母染色体 DNA 中也发现一个 DSB 存在于位点 2（以及位点 1），并且这些 DSB 出现的时间也与质粒中相同。

Nancy Kleckner 等的研究表明，当在 *HIS4* 基因附近插入一个 *LEU2* 基因形成一个减数分裂重组热点时，相似的双链断裂也发生在酵母中。实际两个相距很近的 DSB 一起发生在这个重组热点区。*RAD50* 基因（*rad50S*）发生一个非无义突变（一个不完全阻滞基因活动的突变），也导致由 DSB 产生的一个重建的片段。这个突变很明显阻断了 DSB 下一步的形成，导致 DSB 的积聚。

在 1995 年，Scott Keeney 和 Kleckner 发现在 *rad50S* 突变体内由 DSB 产生的 5′端共价结合了一种蛋白质。根据这一行为提出的一个有吸引力的假说认为，产生的 DSB 将使具有催化功能的蛋白质很快与 DNA 分离，但是在该突变体内催化蛋白仍然结合在它所催化产生的 DNA 末端。如果这是正确的，那么鉴定蛋白质是否结合在双链 DNA 断裂末端就可以鉴别出催化产生双链 DNA 断裂的内切核酸酶的种类。

为此，Kleckner 等开展了鉴定共价结合 DSB 的蛋白质的实验。他们从 *rad50S* 细胞中分离细胞核。由于这些细胞有蛋白质-DSB 复合体的积累，因此成为提供足量 DSB 结合蛋白质的来源。为纯化结合蛋白质，Kleckner 等用了一个两步的鉴定程序：第一，提取核并用胍和去垢剂变性蛋白质，接着用 CsCl 梯度超速离心纯化 DNA 和 DNA-蛋白质复合体。在这种变性条件下，结合 DNA 的任何蛋白质都应该发生共价结合。第二，通过玻璃纤维膜过滤混合物，它只允许纯 DNA 通过，DNA-蛋白质复合体结合到膜上。这样，在滤膜上就会富集共价结合的 DNA-蛋白质复合体。Kleckner 等用核酸酶消化复合体中的 DNA，对释放的蛋白质进行 SDS-PAGE 电泳。他们观察到许多带，其中 2 条带在 *rad50* 细胞中出现，但并不在阻断 DSB 形成的 SPO*11*Δ 中出现。这两条相同的带也出现在制备性阶段以及中间试验准备阶段中，它们的出现与否依赖于 DNA 蛋白质复合体的核酸酶处理。

接着，Kleckner 等从制备性凝胶上切下两条带（M_r＝34kDa 和 45kDa）并用胰酶消化，测序其中的一些胰酶肽。从这些肽链产生的短的蛋白质序列中得出了相应的 DNA 序列。由于酵母染色体基因组的序列已知，很容易对这些短的序列与相关的基因进行比对。结果表明 45kDa 的蛋白质是 *spo11* 基因相应的产物 Spo11，而这些 34kDa 的蛋白质对应于 5 种不同蛋

白质的混合物，其中包括两种核糖体蛋白质。

由于 Spo11 是减数分裂必需的，因此它是 DSB 结合蛋白质的一个值得关注的候选基因。为了增加这种假设的说服力，Kleckner 等证实 Spo11 特异性的与 DSB 结合，并不是结合大多数的 DNA。为证明这一点，用抗原决定域标签途径，构建了红细胞凝集素蛋白的一个抗原决定域基因编码区与 *spo11* 基因的融合基因。该基因的蛋白质产物（Spo11-HA）与任意 DNA 连接后，都可以与抗红细胞凝集素的抗体发生免疫沉淀反应。

在预实验中，研究人员从包含 *HIS4LEU2* 重组热点的 rad50S 减数分裂细胞中分离 DNA（以及与其共价结合的蛋白质）。利用 *Pst*I 酶切 DNA，电泳分离，用从重组热点制备的探针进行点杂交。图 22.20a 描绘了热点区域，显示减数分裂中 DSB 发生

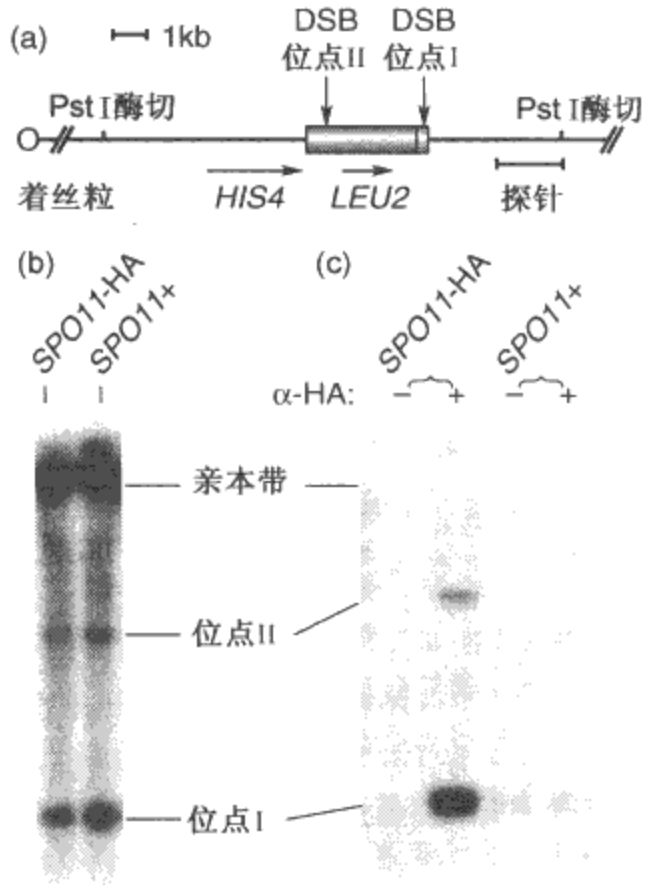


图 22.20　Spo11 与 DSB 片段的结合。（a）重组热区图谱。将带有 LEU2 基因（红）的片段（红和蓝），插入到酵母染色体 III 近 *HIS4* 基因位点。着丝粒（*CENIII*）DSB 位点 I 和 II 的位置，Southern 印迹探针杂交的位点以及在 DSB 两侧两 *Pst*I 位点均分别在图中标示。（b）总 DNA 的 Southern 印迹，*Pst*I 酶切，电泳，转膜，杂交。亲本的 DNA 片段，以及 DSB 产生的亚片段均显示出来。（c）DNA 的 Southern 印迹。*Pst*I 酶切，与抗 HA 抗体免疫沉淀 Spo11-HA-DNA 复合体，与图（b）相同的转膜，杂交。DSB 产生的亚片段比亲本的片段更好地富集了。(*Source*: Keeney, S., C. Giroux and N. Kleckner, Meiosis-specific DNA doublestrand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell* 88 (Feb 1997) f. 3, p. 378. Reprinted by permission of Elsevier Science.)

的两个位点，在 DSB 位点的两侧有两个 *Pst*I 位点，以及 *HIS4LEU2* 下游探针杂交的位置。如果没有 DSB 发生，仅仅可以观察到亲本的 *Pst*I 片段。另一方面，如果 DSB 的确发生了，对应于 DSB 位点 1 和 2 的两个额外的附加小片段会出现。图 22.20b 表明无论是在野生型细胞（SPO11$^+$）还是在 SPO11-HD 细胞内，两个小片段都出现了。

接下来，Kleckner 等检测到 Spo11-HA 特异地结合到 DSB 产生的片段上。他们重复了上述实验，不过，这次他们在用 *Pst*I 酶切 DNA 后，免疫沉淀 Spo11-HA-DNA 复合体。图 22.20c 描述了实验结果。图 22.20 中明显看出有两条较亲本片段小得多的双链 DNA 断裂片段，伴随着 Spo11-HA 发生了免疫沉淀。在缺乏抗-HA 抗体时，不能发生免疫沉淀。在含有野生型 SPO11 基因而没有 HA 标签连接的酵母菌株中也没有发生沉淀。进一步分析表明，在没有 DSB 积累的野生型 *RAD50* 株系，或者在不形成 DSB 的突变体株系中，都不能发生免疫沉淀。

如果 Spo11-HA 仅仅结合到 DNA 上，那么它与亲本 DNA 片段的连接就应该同与由双链 DNA 断裂产生的亚片段的连接是一样的。但是在免疫沉淀中富集的亚片段大约是亲本 DNA 的 600 倍。Spo11 表现为特异地结合 DSB，并且可能是催化产生 DSB 的酶的一部分。此外，Spo11 与其他生物体中已知的一些蛋白质是同源的，包括始祖鸟、分裂酵母和蛔虫。这 4 种蛋白质仅有一个保守的酪氨酸，该酪氨酸很有可能是共价结合 DSB 的催化氨基酸。按照这种情况，该氨基酸应该类似于拓扑异构酶的酪氨酸活性位点（见第 23 章）。在 Spo11 中保守的酪氨酸是具有活性的 tyr-135。因此，可以提出一种如图 22.21 所示的模型，需要 2 分子 Spo11 的参与，其中每 1 分子 Spo11 在稍微偏移的位置上分别作用 DNA 一条链，产生 DSB，并且留下一个瞬时的中间体，这个中间体带有一个 Spo11 分子，其通过活性酪氨酸共价连接到每条链新产生的一个 5′-磷酸末端上。因此，DSB 的形成显然不是一个简单的水解作用产生的，而是要通过一个转酯反应（transesterification），在该反应中，作用基团是酶的酪氨酸残基，而不是水分子。

Spo11 与 DSB 之间的共价连接仅仅是短暂的过渡现象。因此，2 分子的 Spo11 必须通过一定的方式被除去。这个过程可能通过蛋白质-DNA 结合键的直接水解，或者通过内切酶的作用从每一个末端将蛋白质以及一个短的延伸 DNA 除去。

2005 年，Scott Keeney 等证实后一种机制是正确的，如图 22.21 所示。他们构建了一个酵母菌株来表达带有 HA 抗原决定簇标签的 Spo11。接着，

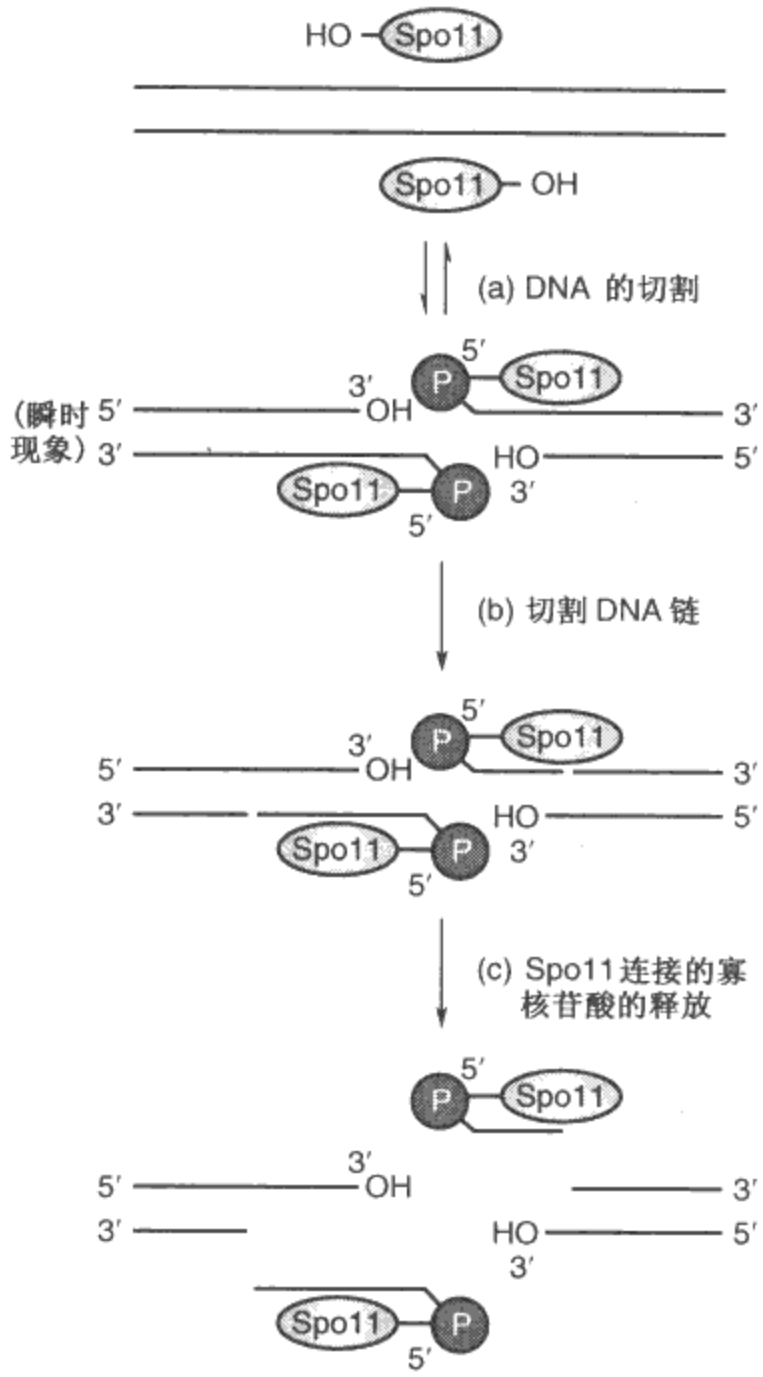


图 22.21　SPO11 参与 DSB 形成的模型。(a) DNA 的切割。两具有活性酪氨酸（以羟基表示）的 Spo11 分子在位置稍有偏移的切点上攻击 DNA 的两条单链，转酯反应使 DNA 链的磷酸二酯键断裂。Spo11 的酪氨酸与被切割的 DNA 5′端上的磷酸发生共价连接。(b) 切割 DNA 链。不对称的切割产生两种不同大小的与 Spo11 连接的寡核苷酸。(c) Spo11 连接的寡核苷酸的释放。如图所示，释放发生在 DNA 末端被切割之前，但有证据表明随后发生释放。

利用 1 个抗 HA 抗体从减数分裂细胞中免疫沉淀 Spo11。为了检测与 Spo11 结合的核苷酸，利用 TdT 和^{32}P 标记的 3′-脱氧腺苷磷酸盐进行免疫沉淀。TdT 加寡核苷能非特异地结合 DNA 的 3′端，而 3′-脱氧腺苷磷酸盐（3′-deoxy adenosine triphosphate）因不提供结合到下一个核苷的 3′-羟基而终止反应。

图 22.22 显示实验结果。由于在没有加细胞提取物和抗体的第一泳道中也可以看到条带，因此可

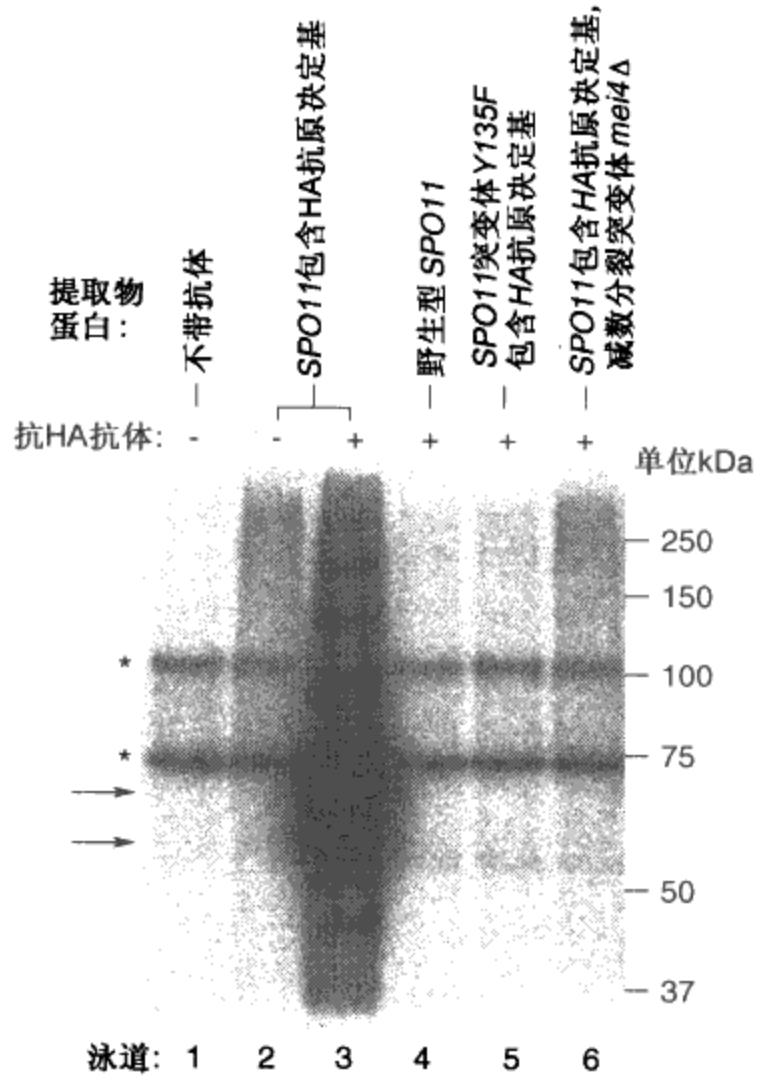


图 22.22　Spo11 与寡核苷酸结合的证据。Keeney 等利用无抗体（泳道 1 和 2）或抗 HA 抗体来免疫沉淀酵母细胞提取物中的蛋白质，如上面所示。用于制备酵母提取物的细胞的基因型也标注在顶部：泳道 2 和 3，*SPO11* 结合到 HA 抗原决定基的编码区域；泳道 4，野生型 *SPO11* 不包括抗原决定基；泳道 5，*SPO11* 突变体 *Y135F*（其中活性酪氨酸位点 Y 转变为苯丙氨酸 F）；泳道 6，减数分裂突变体 *mei4*Δ，没有形成 DSB。Keeney 等末端标记寡核苷酸，结合上附带 TdT 和 alpha-^{32}P免疫沉淀的三磷酸盐。最后，对蛋白质 SDS-PAGE 电泳，放射自显影技术检测标记的蛋白质，星号表明标记的非特异性带，该带甚至在没有酵母提取物或抗体的情况下也会出现。箭头表明 Spo11 特异蛋白仅在 DSB 形成时被标记。（*Source*: Reprinted by permission from Macmillan Publisher Ltd: *Nature* 436, 1053-1057, Thomas Schalch, Sylwia Duda, David F. Sargent and Timothy J. Richmond, "Endonucleolytic processing of covalent protein-linked DNA double-strand breaks," fig. 16, p. 1054 copyright 2005.）

以推测带星号标记的两条带与 DSB 或 Spo11 无关。箭头所指两条带（泳道 3）是 Spo11 特异性带。这些特异性的条带是真实的，在泳道 2 中用无 HA 抗

体的免疫沉淀反应时或 Spo11 没有标记 HA 抗原域时（泳道 4），条带不出现。这些条带的出现与否也依赖于 DSB 的形成，当 Spo11 的催化酪氨酸变成苯丙氨酸（泳道 5）或者 DSB 被 *mei4* 阻断时（泳道 6），条带不出现。

事实上，寡核苷酸标记的 Spo11-HA 出现在两条带中，这表明 Spo11 与寡核苷酸结合产生两种大小不同的带。因此，Keeney 等用蛋白酶从每一条带中消化蛋白质并电泳剩余的寡核苷酸来确定其大小。上面的带为 24～40 个碱基的弥散寡核苷酸带，小的带为 10～15 个碱基的弥散寡核苷酸带。结果证实两种大小的寡核苷酸结合到 Spo11 上，弥散带暗示着寡核苷酸具有不同的长度，或蛋白酶在切断 Spo11 时是异质的，并且允许平均 3 个氨基酸残留仍然结合在寡核苷酸上。通过凝胶分离证明小于 10 个碱基的寡核苷酸不能很好的被阻滞。Keeney 等估计两个寡核苷酸的长度分别是 21～37nt 和小于等于 12nt。对老鼠 DNA 中 DSB 的形成过程进行研究得到了相似的结果，但与老鼠 Spo11 同源部分链接的两组寡核苷酸的大小与酵母中发现的寡核苷酸的大小不同。

图 22.23 表明 Spo11-寡核苷酸复合体出现的时间与 Spo11 切割的 DSB 出现时间的是完全一致的。预测 Spo11-寡核苷酸是 DSB 产生过程的自然产物。此外，大的条带和小的条带的比率也完全符合1∶1，表明两者是同时产生于同一个过程。这些发现得出一些有趣的尝试性的结论。

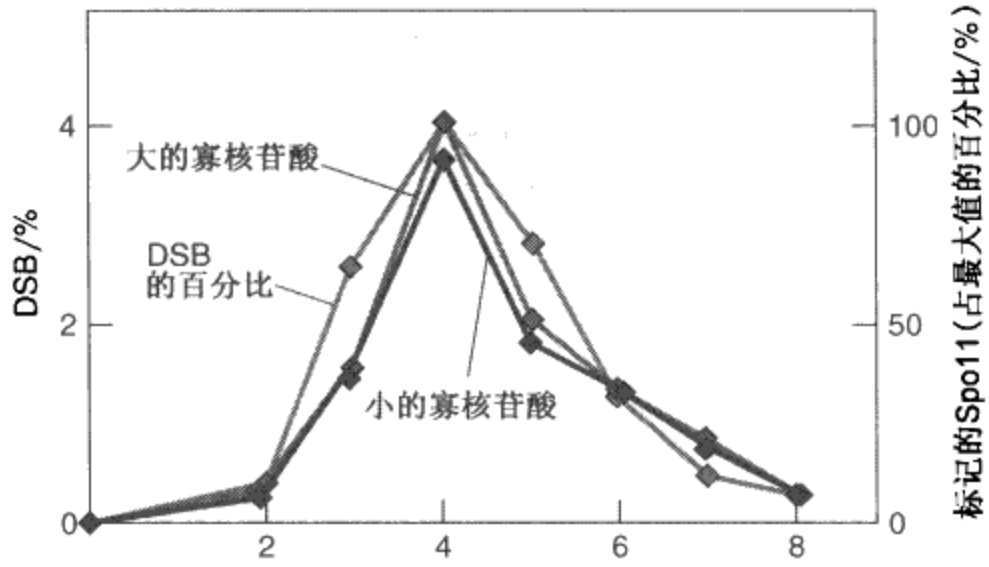


图 22.23 DSB 和 Spo11-寡核苷酸复合体形成以及消失的时间过程。 在酵母 *HIS4LEU2* 重组热点测定的 DSB 占总 DNA 的百分比（绿色）。大的（红色）和小的（蓝色）寡核苷酸连接的 Spo11 的种类依照大种类的最大的百分比划分。（*Source*: Adapted from Neale, M. J. et al., Endonucleolytic processing of covalent protein-linked DNA double strand breaks. *Nature* 436: 1054, fig. 1f, 2005.）

首先，Spo11-寡核苷酸的积累与消失基本反映了 Spo11 切割的 DSB 的积累与消失。相应物种的 Spo11-寡核苷酸的积累与 Spo11 切割的 DSB 的积累是一致的。但没有人能断言，Spo11-寡核苷酸的消失将与 Spo11 切割的 DSB 的消失一致。然而，最简单的模型是 Spo11（结合或不结合寡核苷酸）在切除前从 DSB 上释放（图 22.21c），而这将发生 DSB 消失前，由于 DSB 参与 Holliday 联接体形成。这个模型预测 Spo11-寡核苷酸的消失发生在 Spo11 切割的 DSB 消失之前。这两个现象同时发生的事实可以通过两种途径来解释。第一，Spo11-寡核苷酸的缓慢破坏足以使得同时发生 DSB 的切除而形成 Holliday 联接体。第二，但更有趣的是可能 Spo11-寡核苷酸直到 DSB 切除后才开始降解，因为 Spo11-寡核苷酸直到那时才被释放。图 22.24 阐明了这个概念。

图 22.23 的第二个耐人寻味的结果是产生了两种数量相同、大小不等的 Spo11-寡核苷酸连接体。这暗示着大的寡核苷酸来源于 DSB 的一个末端，而小的寡核苷酸来源于 DSB 的另一端。这意味着 DSB 在选择 3′游离端侵入另一个 DNA 双螺旋来起始 Holliday 联接体的形成之前具有固有的不对称性。Keeney 等设想了一个与图 22.24 相似的模型。两条链切割的不对称性导致了切除后同 Spo11 连接的寡核苷酸碱基互补配对的 3′游离末端长度的不对称性。图 22.24 右边的链碱基配对不紧密，将结合两个重组酶（Rad51 或者 Dmc1）中的一个，而另一个链将结合另一个重组酶。这种不对称性将支配着 3′游离末端侵入同源双链来起始 Holliday 联接体的形成。这种情况，正如图 22.24 右边所示。

下一节将在切除 DSB 末端的过程中看到一个包含 *Rad50* 和 *Mre11* 的复合体，下列证据表明复合体也包含内切核酸酶，切割靠近 DSB 的 DNA，导致了 Spo11-寡核苷酸的释放。第一，Mre11 具有必需的内切核酸酶活性；第二，*RAD50* 和 *MRE11* 的突变基因阻断了 Spo11 从 DSB 末端的移除；第三，结合于 Spo11 上的寡核苷酸具有 3′-羟基基团，与 Mre11 内切核酸酶作用的机制一致。

Spo11 基因在包括酵母、植物、动物的真核生物中是高度保守的。双链断裂模型起始重组似乎也是保守的。Kim McKim 和 Aki Hayashi-Hagihara 等进行了与 Kleckner 等相似的实验，研究结果支持这个结论，在果蝇中寻找阻断基因转换的突变体，并在 1998 年，报道 *mei-W68* 突变基因具有这样的表型，而 *mei-W68* 是果蝇中与酵母 *Spo11* 基因同源的

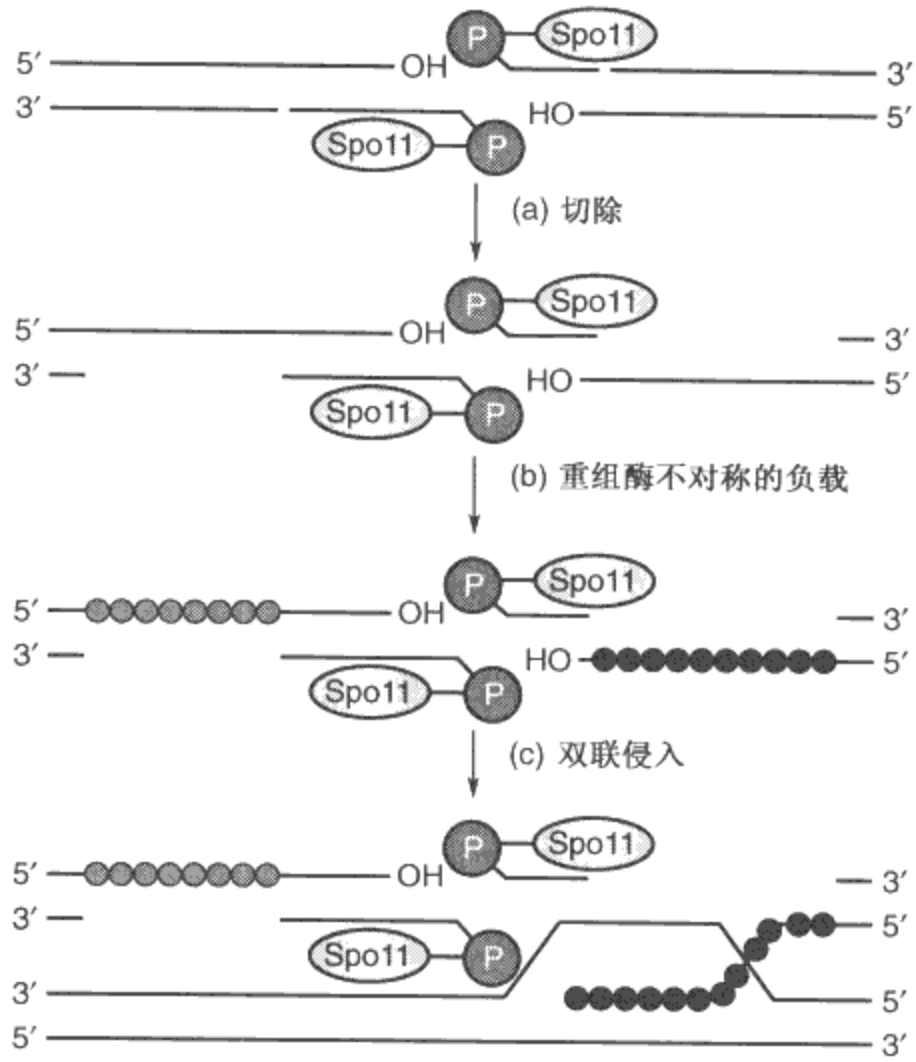


图 22.24　Spo11-寡核苷酸释放前 DSB 末端切除的模型。（a）利用前一步（图 22.21）产生的缺口对两条链进行切除。（b）两个重组酶（Rad51 和 Dmc1）不对称地负载到新产生的单链区域，一蛋白质（蓝色）包裹一条链，另一蛋白质（橘色）包裹另一条链。从这里不知道那一个蛋白质启动链拆分，颜色是随意的。（c）其中一个蛋白质（蓝色）标注被紧密包裹的 3′自由端，入侵一个同源双螺旋，起始 Holliday 联接体的形成。

基因。有趣的是，*mei-W68* 突变影响体细胞和减数分裂细胞的基因转换。因此，与酵母中仅在减数分裂重组时需要 Spo11 的情况不同的是在果蝇中 *mei-W68* 对减数分裂和体细胞重组都是需要的。可见，在果蝇中由 *mei-W68* 诱导的 DSB 对减数分裂和体细胞的重组都是必需的。

小结　酵母减数分裂重组模型起始于一个双链 DNA 的断裂。DSB 可以直接在质粒 DNA 或染色体 DNA 中被观察到。DSB 在 *rad50S* 突变体中积累，其中 Spo11 可以与 DSB 末端共价结合。2 个 Spo11 分子共同作用在间隔邻近的两个位点切割 DNA 双链产生 DSB。这种切割涉及在 2 个 Spo11 分子上的酪氨酸活性位点发生的交换反应，导致 2 个 Spo11 分子和新形成 DSB 之间共价连接。接着 Spo11 与寡核苷酸（12～37nt 的寡核苷酸）的复合体从 DSB 处释放。Spo11-寡核苷酸复合体在 DSB 切除后释放。Spo11 的这种切割是不对称的，在 DSB 的一边产生比另一边长的游离的 3′端。这样可以形成一个游离的末端侵入同源 DNA 双链，起始 Holliday 联接体的形成。

DSB 处单链末端的生成

一旦 Spo11 产生一个 DSB，新的 5′端将被消化产生可以侵入另一 DNA 双螺旋游离的 3′端。1989 年，Szostak 等研究 DSB 产生的 DNA 末端结构时首次发现单链末端。他们用可以完整的保留双链 DNA 但特异降解单链末端 DNA 的 S1 酶消化 DNA。S1 酶不影响带的强弱，但减少了 3 个 DNA 链的长度。这一结果与图 22.18 中的模型所预测的一致。DSB 发生后，外切核酸酶在切点处降解两个 5′端，形成了对 S1 核酸酶敏感的单链 DNA，S1 消化后将会减少 DNA 片段的长度。

Kleckner 等也发现了 DSB 处一条链降解或切除的证据。他们发现野生型细胞积累的片段用凝胶电泳分析产生弥散带型，这是切割末端后形成不同大小的片段的结果。而在突变体 Rad50S 中，因为 DSB 末端的切割被阻断从而得到离散带型。

RAD50、*MRE11* 和 *COM1/SAE2* 的突变阻断了 DSB 的切割。实际上，*RAD50* 和 *MRE11* 的无效等位基因一起阻断了 DSB 的形成。所以只有这些基因的几个有效等位基因允许 DSB 形成，但阻止 DSB 的切除。可见这两种基因的产物对 DSB 的形成和切割都是必要的。这些基因产物功能的证据来源于与 *E. coli* 的比较。*E. coli* 蛋白质 SbcC 和 SbcD 分别与酵母 Rad50 和 Mre11 是同源蛋白。此外，两种细菌蛋白质发挥作用时形成 SbcC/SbcD 二聚体，它具有双链 DNA 的外切核酸酶活性。这些发现表明 Rad50 和 Mre11 在酵母中一起切割酵母 DSB 中的 5′端。

小结 减数分裂重组过程中，在 DSB 形成后，断口处的 5′端被 5′→3′外切核酸酶切割降解，产生能侵入 DNA 双螺旋的 3′突出端。Rad50 和 Mre11 可能相互协作来完成这一过程。

22.4 基因转换

当真菌（如红色面包霉菌）形成孢子时，两个单倍体核融合，产生一个二倍体核。经过减数分裂形成 4 个单倍体核。这些核经有丝分裂产生 8 个单倍体核，其中每一个都成为一个独立的孢子。总而言之，如果在某一个原始核的假定位点上含有一个等位基因 A，而另一个核在同一个位点含有等位基因 a，那么在孢子中这些等位基因的数量应该是相当的，即 4 个 A 基因和 4 个 a 基因。很难想象其他类型的结果存在（如 5 个 A 和 3 个 a），因为这需要一个 a 转换为 A。事实上，这种变异率大约为 0.1%，不同的真菌会有不同的变异率。这种现象被称为基因转换（gene conversion）。我们在重组的主题下讨论基因转换，是因为这两个过程是相关的。

图 22.18 所展示的减数分裂重组的机制揭示了减数分裂中基因转换的机制。图 22.25 描述了基于 *N. crassa* 的一个假设。一个细胞核发生 DNA 复制后，会含有 4 条染色单体。理论上，2 条染色单体具有 A 等位基因，两条染色单体具有 a 等位基因。但是，如果链交换和分支端化已经发生，随后的拆分将产生两条带有异源补丁的染色体，如图 22.25。异源双链区域在等位基因 A 和 a 所在的位置仅有单个碱基不同，所以染色体中有一条链携带这个等位基因，另一条链携带另一个等位基因。如果立即发生 DNA 复制，将会产生 2 个 A 链和 2 个 a 链。然而，如果在复制前，1 个或者 2 个异源双链在酶的作用下进行碱基错配修复。正如图 22.25 所示，仅有上面的异源双链被修复，有 1 个 a 被转换成 A。结果是 3 条链带有 A 等位基因，仅有 1 条链带有 a 等位基因。DNA 复制后，将产生 3 个 A 双螺旋和 1 个 a 双螺旋。加上未形成异源双链染色体的 2 个 A 和 a 双螺旋，最后的结果是 5 个 A 双螺旋和 3 个 a 双螺旋。

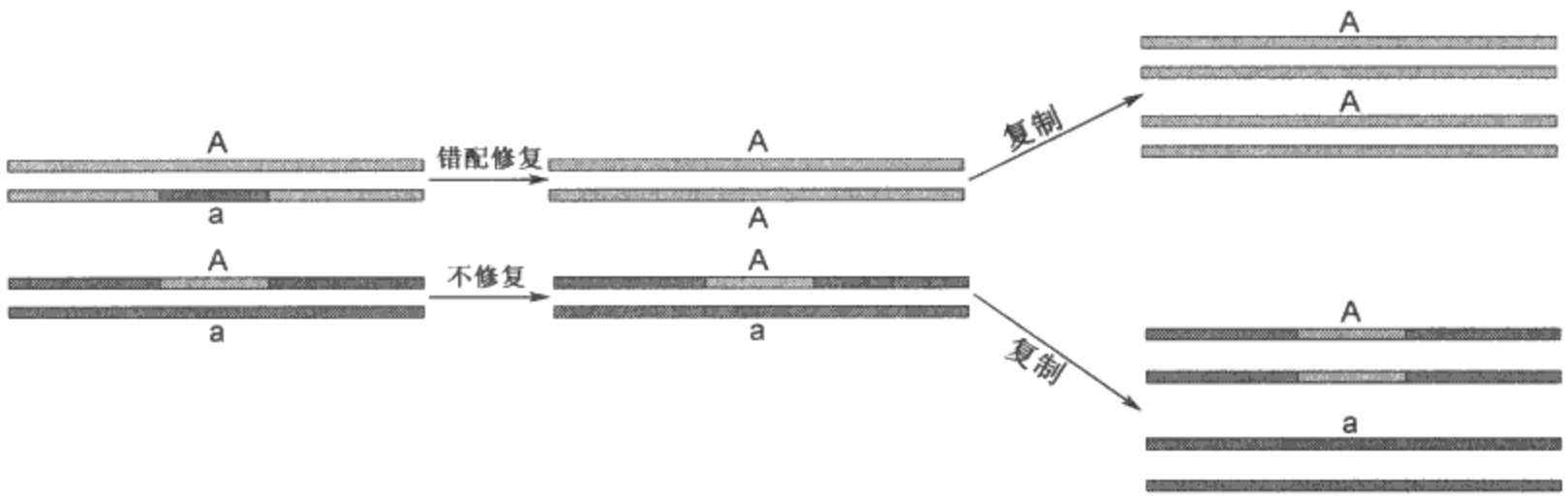


图 22.25 脉孢菌属孢子形成过程中的基因转换模型。在孢子形成中，伴有分支端化过程的重组交换事件会产生两条带“补丁”的异源双链 DNA。异源双链 DNA 等位区域 A（蓝色）/a（红色）之间仅相差一个碱基。另外产生具有纯合 A 和纯合 a 基因的两条子代 DNA（图中未显示）。中间图所示，上面一条异源双链经错配修复，a 被转换为 A，底部的异源双链没有经过修复。修复的 DNA 复制产生两个含 A 基因的双链 DNA 分子。未修复的 DNA 复制产生含有 1 个 A 基因和 1 个 a 基因的双链 DNA。右侧图中子代双链 DNA 含有 3 个 A 基因和 1 个 a 基因，加上未显示的 2 条 DNA 复制产生 2 个 a 和 2 个 A 的双链 DNA。最后在子代的双螺旋中有 5 个 A 和 3 个 a，而不是 4 个 A 和 4 个 a。

图 22.26 描述另一种基因转换发生的途径，这次没有错配修复。从图 22.18 步骤 c 开始，即从链侵入和 D 环形成开始。等位基因 A 和 a 的差异区在顶图中显示出来。蓝色代表 A，红色代表 a。图 22.26 与图 22.18 不同之处在于在修复合成之前，侵入链受部分切除的影响，D 环带有部分塌陷。这种切割使得修复合成能获得更长的伸展，从而侵入链中更多的 A 将转换成 a。转换准确发生在等位基因 A 和 a 差异的区域。分支端化和拆分后（交换或者没有交换的拆分），我们得到的 4 条 DNA 链都具有 a 等位基因，而最初只有两条链是这种情况。基因转换已经发生。

基因转换并不仅仅局限于减数分裂事件。它也是 Baker 酵母交配类型转换的机制。该基因转换事件包含 MAT 位点的两种基因型的瞬时互作，以及随后的一个基因的序列转换为另一个基因的序列。

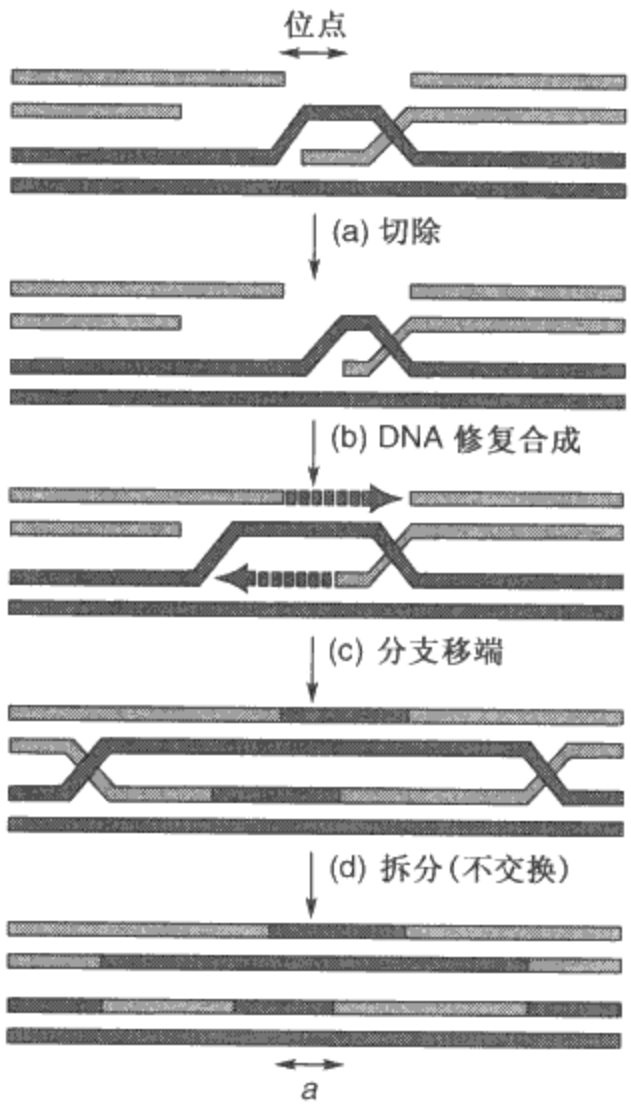


图 22.26　无错配修复的基因转换模型。该图是从图 22.18 所示的 DSB 重组图解的中间部分，即从链侵入之后开始。(a) 此次，侵入链被部分切断，导致 D 环部分塌陷。(b) 由于链的切割，DNA 修复合成更加广泛。将生成一个 4 条 DNA 链均为等位基因 a 的区域（在顶部和底部显示）。(c) 和 (d) 分支端化和拆分均没有改变 4 条 DNA 链上等位基因的 A 和 a 种类，所有的 DNA 链均含有等位基因 a。该过程将一个含有等位基因 A 转变为等位基因 a。

小结　当两条相似但不一致的 DNA 序列互作时，就可能存在基因转换——一个 DNA 序列转换成另一个。参与基因转换的序列可能是等位基因，若在减数分裂中，也可以是非等位基因，如决定酵母交配类型的 MAT 基因。

总结

对于生命体而言，同源重组是非常重要的。在真核生物减数分裂中，保证了同源染色体的正确分离。同样可以使子代获得重组的亲本基因。在所有的生物体中，同源重组可以进行 DNA 修复。

在 *E. coli* 中，同源重组 RecBCD 途径起始于双链断裂产生的一个单链末端侵入另一双链 DNA。通过优先切割一个特异 *Chi* 序列处，由 RecBCD 的核酸酶和解旋酶活性产生一个游离末端。侵入链由 RecA 和 SSB 包裹。RecA 帮助侵入链与同源 DNA 在同源区域配对，并形成 D 环。SSB 很明显地通过解开二级结构，以及阻止 RecA 被任何二级结构所捕获从而抑制重组过程后续的链的交换，而得以加速重组过程。接下来可能是在 RecBCD 的帮助下 D 环链产生缺口，导致 Holliday 联接体中间体的形成。RecA-SSB 解旋酶催化分支端化，将 Holliday 联接体的交叉带到要拆分的位点。最后 RuvC 可以通过切割两条链来拆分 Holliday 联接体，产生两条带有异源补丁的 DNA 或双交换重组体 DNA。

酵母减数分裂重组起始于一个双链断裂。2 个分子的 Spo11 相互作用通过在距离很近的位点切割双链产生 DSB。切割过程中包括 2 分子 Spo11 的酪氨酸活性位点的酯交换反应，导致 2 分子 Spo11 的共价连接和新 DSB 的形成。随后在 DSB 处，从 Spo11 与 12～37nt 寡核苷酸的复合体中释放 Spo11。Spo11-寡核苷酸的释放也可能是在 DSB 切除之后。减数分裂重组中 DSB 形成后，在断裂处的 5′端，按 5′→3′方向降解 DNA。*Rad50* 和 *Mre11* 可能一起作用来完成切除。接着，新产生的 3′突出端侵入其他的 DNA 双螺旋，产生 D 环。DNA 修复合成和分支端化产生可以被拆分成交换或非交换重组体的两 Holliday 联接体。

当两个相似但不一致的 DNA 序列进行互作时，就可能发生基因转换——一个序列转换成另外的序列。参与基因转换的序列可以是减数分裂中的等位基因，也可以是非等位基因，如决定酵母交配型的 MAT 基因。

复习题

1. 列出并解释 RecA 参与的同源重组的 3 个步骤。

2. 证实 RecA 包裹单链 DNA 的证据是什么？RecA 与单链 DNA 互作中 SSB 起什么作用？

3. 描述证实在重组开始阶段的联会中需要 RecA 的实验并给出该实验的结果。

4. 怎样证实你通过电镜观察所看到的联会是真实的联会，而不是真实的碱基配对？

5. 描述 RecBCD 在 *Chi* 位点附近切割 DNA 的实验以及结果。怎样证明缺口是 RecBCD 引起的，而不是污染物造成的？

6. 说明如何利用凝胶阻滞实验来证实 RecA 在高浓度下可以与 Holliday 联接体结合，而 RecB 自身不能结合，以及 RecA 和 RecB 在很低浓度条件下

的协作可以结合到 Holliday 联接体上。实验中戊二醛的作用是什么?

7. 绘制一张 RuvAB-Holliday 联接体复合体图，包括将要发生分支端化的 Holliday 联接体的交叉结构，RuvB 环，但不包含 RuvA 四聚体。

8. 描述 RuvC 可以拆分 Holliday 联接体的实验及结果。

9. RuvA、RuvB、RuvC 结合到一起后与 Holliday 联接体结合成复合体的证据是什么?

10. 列举酵母减数分裂重组的一个模型。

11. 描述证明酵母减数分裂重组中存在 DSB 的实验以及结果。

12. 描述证明酵母减数分裂重组中 Spo11 共价结合到 DSB 的实验以及结果。

13. 描述说明 Spo11 与两组大小不同的寡核苷酸结合的实验以及结果。

14. 由 Spo11 结合两组不同大小的寡核苷酸，以及消失和出现的时间所阐明的酵母减数分裂重组的机制是什么，并利用图示说明答案。

15. 列举减数分裂基因转换的一个模型。

分析题

1. 画出 Holliday 联接体的图表，根据图表举例说明：

a. 向右分支端化，接着拆分产生一个短的异源双链，或者交换的重组体 DNA;

b. 向左分支端化，接着拆分产生一个短的异源双链，或者交换的重组体 DNA。

2. 绘制一张 *E. coli* 细胞具有下列突变体基因时 RecBCD 途径的重组产物图：

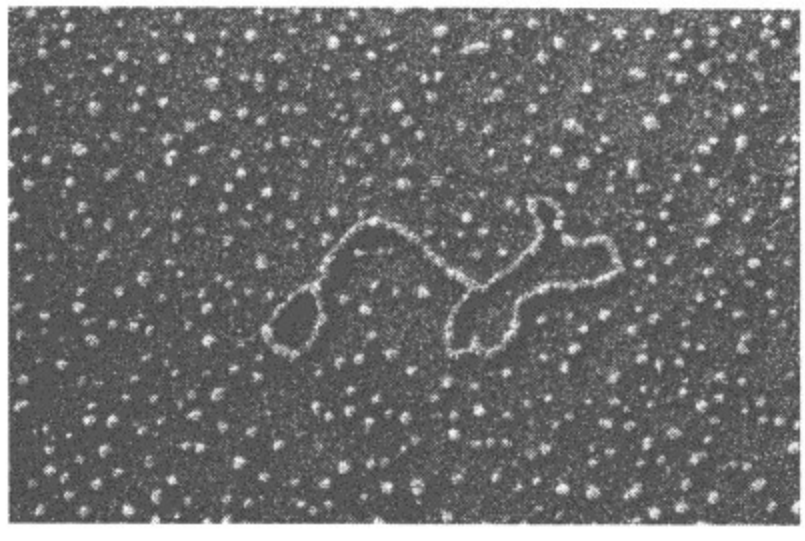
a. *recA*　b. *recB*　c. *ruvA*　d. *ruvB*　e. *ruvC*

3. 说明如何用 DNA 酶足迹法来证实 RuvA 结合到 Holliday 联接体的中心，而 RuvB 结合到分支端化相关方向的上游。

4. 如果两异源双链都经过错配修复修改为 A/A，那么图 22.25 描述的基因转换最终会有什么样的等位基因混合比例? 如果一条异源双链被修改为 A/A，另一条修改为 a/a，情况又如何?

翻译　张姿丽　校译　郑用琏　刘智捷

第23章 转　座



遗传工程转座子 DNA 的假彩色显微图（放大倍数为 140 000）
（*Source*：© *Stanley N. Cohen/Photo Researchers, Inc.*）

我们知道有机体的 DNA 从生命的开始到结束并非保持绝对不变。第 20 章中讲到 DNA 会受到损伤并得以修复，也能发生修复后突变。在第 22 章中已了解 DNA 可以通过同源重组将基因以新的组合方式连接在一起。DNA 还能发生位点特异性重组，它比同源重组所要求的同源序列要短，这种类型的重组常发生在特定的 DNA 序列之间，因此被命名为“位点特异性重组”。一个经典的例子是 λ 噬菌体 DNA 插入到宿主 *E. coli* 的 DNA 中，然后又从 *E. coli* DNA 上切离下来。相比而言，转座很少需要重组的 DNA 序列之间有同源性，因而也不是位点特异性的。转座将是本章的主题。

23.1 细菌转座子

转座，即一种转座因子或转座子，从 DNA 的一个位点移到另一个位点。20 世纪 40 年代，Barbara McClintock 在研究玉米遗传时发现了转座子。随后，在从细菌到人的各种有机体内均发现了转座子的存在。本章首先介绍细菌转座子。

细菌转座子的发现

20 世纪 60 年代晚期，James Shapiro 等对行为异常的噬菌体突变体的研究，为细菌转座子的发现奠定了基础。例如，这些突变体不像点突变那样容易发生回复突变，而且突变基因含有一长串额外的 DNA 序列。Shapiro 利用 λ 噬菌体裂解 *E. coli* 时偶尔会携带宿主的一段 DNA，并用这段“外源”DNA 整合到自身基因组的事实来说明突变体的异常表现。让 λ 噬菌体携带一段野生型的 *E. coli* 半乳糖苷酶基因（gal^+）或与之对应的突变型基因（gal^-），然后测定混有宿主 DNA 的重组 λDNA 分子的大小（第 20 章）。用氯化铯密度梯度离心方法测定两种类型噬菌体的密度，从而间接测定重组 DNA 分子的大小。噬菌体的外壳由同体积的蛋白质组成，由于 DNA 的密度比蛋白质的密度大，所以如果噬菌体的 DNA 含量越高，其密度也越大。结果显示携带 gal^- 基因的密度比携带野生型 gal^+ 基因的密度大，说明前者含有更多的 DNA。最简单的解释是外源 DNA 插入到 gal^+ 基因中使其失活。后来的实验的确揭示出在突变的 gal^- 基因中插入了一段在野生型基因中检测不到的 800～1400bp 的 DNA 片段。在极少数情况下，当突变体发生回复突变时，它们也丢失额外的 DNA。这些插入到一个基因中而使其失活的额外 DNA 是首先在细菌中发现的转座子，被称为插入序列（IS）。

插入序列：最简单的细菌转座子

细菌插入序列中仅含转座所必需的因子，包括转座子的末端反向重复序列以及编码转座酶的基因。

由于插入序列的末端是反向重复序列，如果一端是 5′-ACCGTAG，那么另一端则是反向互补的 CTACGGT-3′。这里给出假设的反向重复序列来说明这一特点，典型的插入序列含有 15～25bp 的反向

重复序列。例如，IS1 的反向重复序列有 23bp，大的转座子的反向重复序列可以长达上百个碱基对。

Stanley Cohen 用一个巧妙的实验证实了转座子末端具有反向重复序列。如图 23.1 所示。先给出一个结构如图 23.1a 左侧所示的含有转座子的质粒。这个质粒与转座子的末端反向重复序列连接。如果转座子末端确实有反向重复序列，那么可将重组质粒的双链分开，每一条链上的反向重复可以相互配对形成如图 23.1a 右侧所示的茎环结构。茎部是反向重复序列组成的双链 DNA，环是反向重复序列之外其他的单链 DNA。图 23.2b 的电镜图像证实了这种预期的茎环结构。

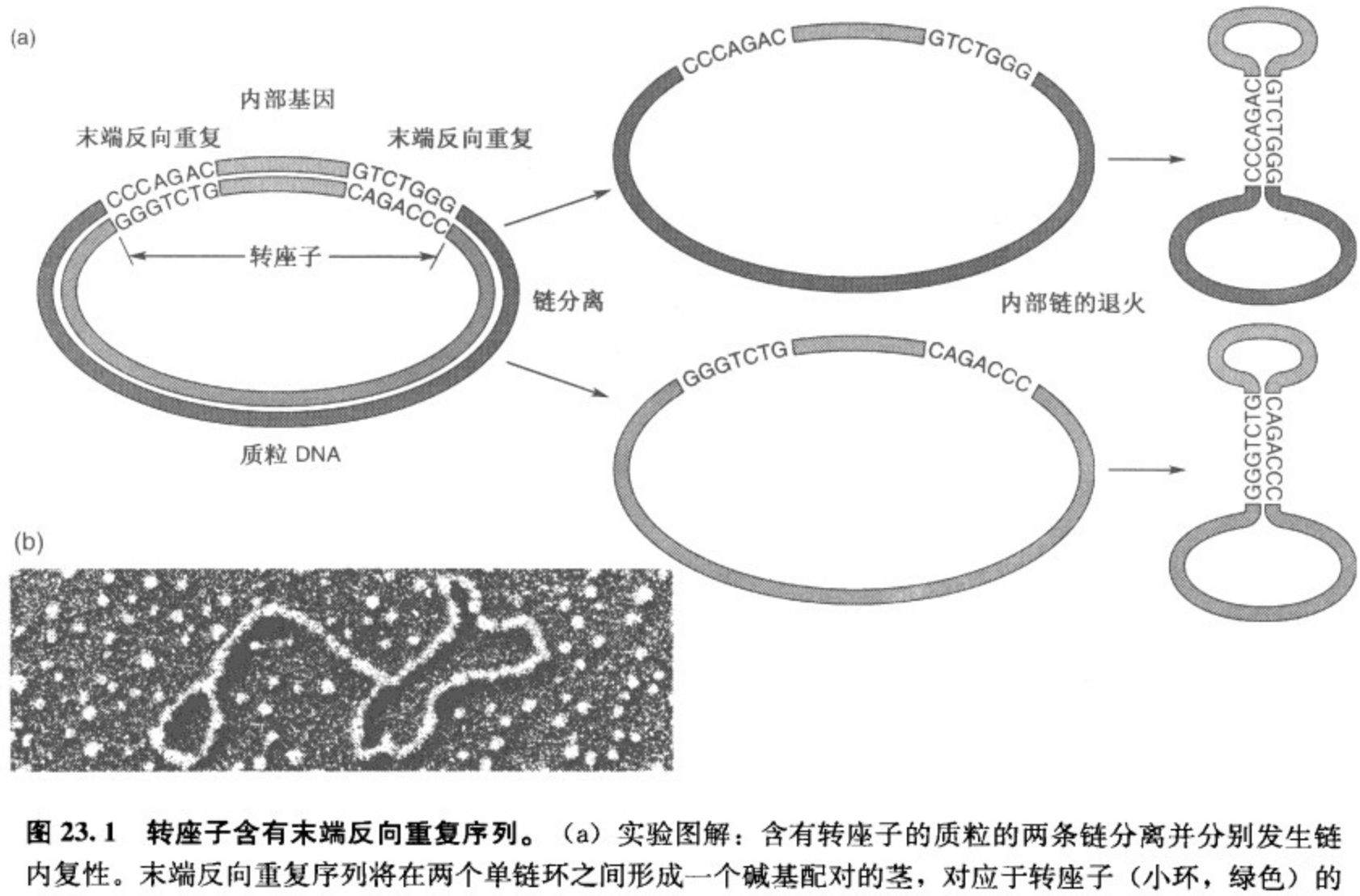


图 23.1　转座子含有末端反向重复序列。（a）实验图解：含有转座子的质粒的两条链分离并分别发生链内复性。末端反向重复序列将在两个单链环之间形成一个碱基配对的茎，对应于转座子（小环，绿色）的内部基因和宿主质粒（大环，紫色和粉色）。（b）实验结果：DNA 用重金属处理成像后在电子显微镜下检测，显示出环-茎-环结构。茎部有上百个碱基对，证明这个转座子中的反向重复序列比在（a）图显示的 7bp 要长得多。（*Source*：（b）Courtesy Stanley N. Cohen，Stanford University.）

插入序列的主体部分至少编码两个催化转座的蛋白质，这些蛋白质统称为转座酶（transposase）。在本章后文中将介绍其作用机制。若插入序列的这个区域发生突变将导致其不能转座，所以这些蛋白质是转座必需的。

和其他复杂转座子一样，插入序列的另一个特征是，在转座子两外侧的靶 DNA 中出现一小段正向重复序列。这段正向重复序列在转座子插入之前是没有的，它产生于转座子插入的过程中，这就告诉我们转座酶剪切靶 DNA 时采用错位剪切方式，即不是正好在两条链的对等位点处切割。图 23.2 显示了靶 DNA 的两条链如何在插入位点错位切割后形成正向重复（DR）。形成的正向重复的长度取决于在靶 DNA 链上两个切点的间距。这个距离反过来又取决于插入序列本身的特性。IS1 的转座酶形成 9bp 的切点，因此产生的正向重复长 9bp。

小结　插入序列是最简单的细菌转座子，仅包含满足自身转座需要的因子：末端短的反向重复序列和由至少两个基因编码的转座酶。复制靶 DNA 的一段短序列，转座后插入序列侧翼各有一个拷贝。

更复杂的转座子

插入序列和其他转座子有时候也被称为“自私 DNA”，即它们利用寄主进行复制却没有给寄主提供任何有用的东西。但有些转座子的确也会给寄主带来有价值的基因，最常见的是抗生素抗性基因。这不仅对细菌寄主有利，对分子生物学家同样有用，因为有了这些基因，可以更加容易地跟踪转座子。

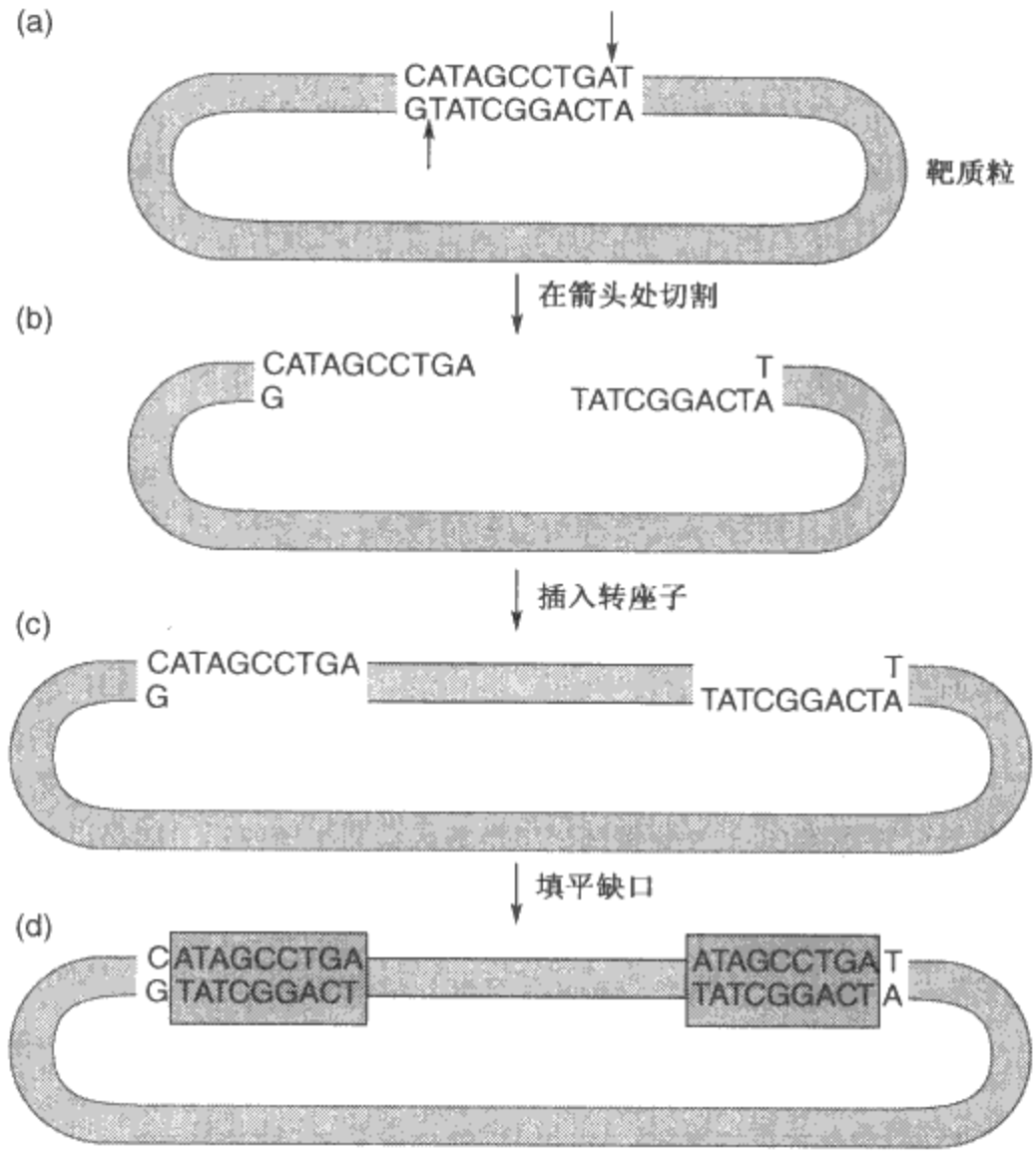


图 23.2　转座子两侧的寄主 DNA 正向重复序列的形成。（a）箭头显示的是寄主 DNA 双链错位切割切口的位置，每边 9bp；（b）切开后；（c）转座子（黄色）连接到寄主 DNA 的每条链的末端，留下两个 9bp 的缺口；（d）缺口补齐后，在转座子的两个末端形成了寄主 DNA 的 9bp 重复（粉色框）。

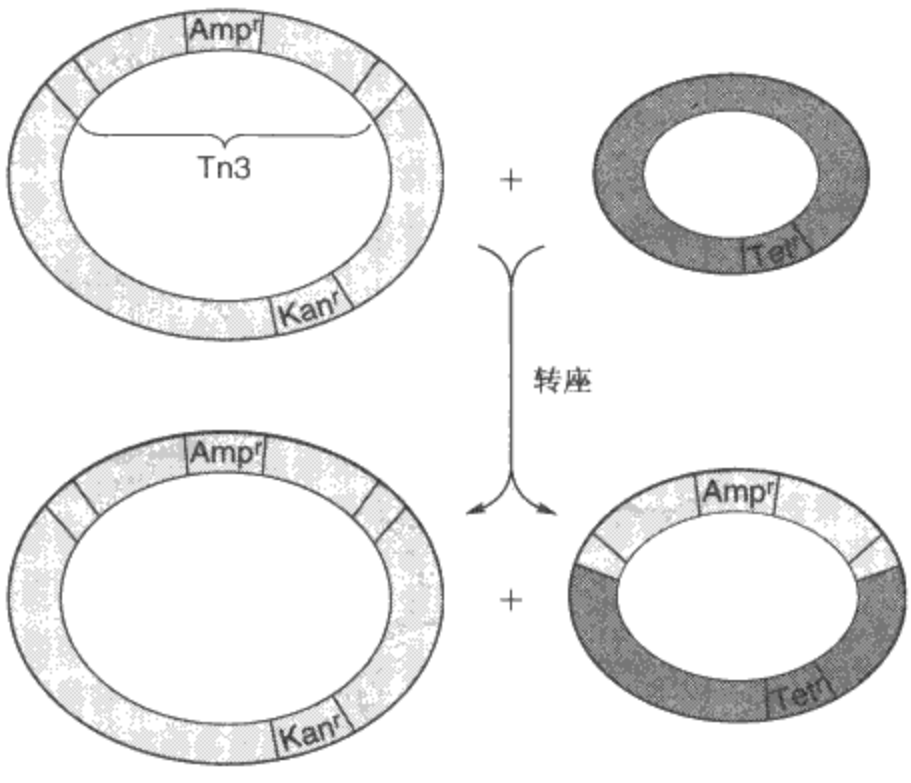


图 23.3　利用抗生素抗性基因追踪转座。开始有两个质粒：大的（蓝色）编码卡那霉素抗性（*Kan*^{r}）并带有编码氨苄青霉素抗性基因（*Amp*^{r}）的转座子 Tn3（黄色）；小的（绿色）编码四环素抗性基因（*Tet*^{r}）。转座后，小段质粒同时携带有 *Tet*^{r} 和 *Amp*^{r} 基因。

如图 23.3 所示，首先有一个供体质粒，其上携带有卡那霉素抗性基因（*Kan*^{r}）和含有氨苄青霉素抗性基因（*Amp*^{r}）的转座子（Tn3）；另外，我们有一个携带有四环素抗性基因（*Tet*^{r}）的靶质粒。转座后，Tn3 通过复制将一个拷贝移动到靶质粒上。现在靶质粒同时具有四环素和氨苄青霉素抗性，由于这个特性，我们可以通过将靶质粒转化到抗生素敏感性的细菌中，并将这些细菌置于含有这两种抗生素的环境中生长来检测转座子。如果细菌能存活，它们必须含有两种抗生素抗性基因，由此可证明 Tn3 必然转座到靶质粒中了。

转座的机制

由于转座子具有从一个地方移动到另一个地方的能力，所以有时候被称作“跳跃基因”。但是这个称呼有些误导，因为“跳跃基因”似乎暗示了 DNA 经常离开一个位点然后跳跃到另一个位点。以这种方式发生的转座，被称为非复制式转座（nonreplicative transposition）（或“剪切-粘贴”），因为初始 DNA 的两条链一起从一个位点移动到另一个位点而没有复制。然而，转座经常包括 DNA 的复制，所以转座子的一个拷贝会保留在原位而另一个拷贝插入到新的位点。这种方式被称为复制式转座（replicative transposition）（或“复制-粘贴”），因为转座子以这种方式移动时也复制了自身。下面我们讨论这两种转座是如何发生的。

Tn3 的复制式转座　Tn3 的结构如图 23.4 所示，这是一种转座机制研究比较清楚的转座子。除了编码钝化氨苄青霉素的 β-内酰胺酶的 *bla* 基因以外，Tn3 含有两个对转座起作用的基因。Tn3 的转座分两步进行，每个步骤都需要 Tn3 的一种基因产物。图 23.5 简要说明了转座发生步骤。我们选用两个质粒开始研究，即一个含有 Tn3 的供体和一个 Tn3 的受体。第一步，伴随着 Tn3 的复制过程，两个质粒融合，通过两个 Tn3 拷贝的连接形成一个共联体（cointegrate）。这一步需要由 Tn3 转座酶基因 *tnpA* 的产物催化在两个质粒间发生重组。图 23.6 显示了所有 4 条 DNA 链如何相互作用实现转座形成共联体的详细过程。图 23.5 和 23.6 图说明在两个质粒间发生转座，但是供体和受体 DNA 可以是其他种类的 DNA，包括噬菌体 DNA 或者细菌自身的染色体。

Tn3 转座的第二步是共联体的拆分，它将一个共联体分解为两个独立的质粒，各自携带一个 Tn3 的拷贝。这一步是发生在 Tn3 自身的同源位点 *res* 间的重组事件，由分解酶基因 *tnpR* 的产物催化。许多证据都证明了 Tn3 转座是由这两个步骤构成的。例如，*tnpR* 基因的突变体不能拆分共联体，因此导致转座的最终产物是共联体形式。这就证实了共联体通常是反应的中间产物。再如，即使 *tnpR* 基因有缺陷，如果有另一个 DNA 分子，比如寄主染色体或其他质粒可以提供有功能的 *tnpR* 基因，共联体仍然可以被拆分。

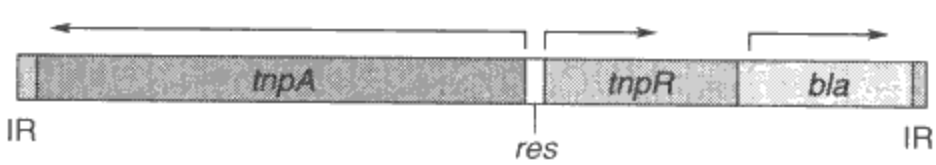


图 23.4　Tn3 的结构。转座必须要 *tnpA* 和 *tnpR* 基因；*res* 是在转座的拆分步骤中发生重组的位点；*bla* 基因编码 β-内酰胺酶，可以让细菌产生氨苄青霉素抗性。这个基因也被称为 Amp^r。在每个末端都有反向重复序列（IR）。箭头所指的是每个基因的转录方向。

非复制式转座　图 23.5 和 23.6 图解说明了复制式转座的机制，但是转座并不总是以这样的方式发生。有些转座子（如 Tn10）采取非复制方式离开供体 DNA，转移到靶 DNA 上。这是如何发生的呢？非复制式转座在起始阶段可能和复制式转座一样，切开缺口，连接供体与受体的靶 DNA，但随后的过程就不一样了（图 23.7）。在转座子的另一边，供体 DNA 形成新的切口，而不是在转座过程中进行复制。这样就释放出产生缺口的供体 DNA，转座子却连接在受体 DNA 上。受体 DNA 的剩余切口可以封闭、产生一个将转座子整合到受体中的重组 DNA 分子。供体 DNA 有一个双链缺口，所以它可能丢失，或者如图 23.7 所示，缺口可以被修复。

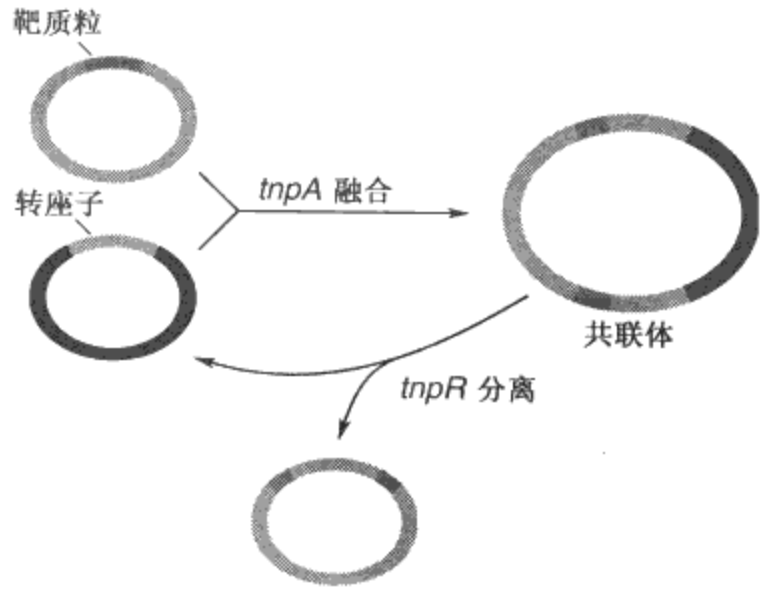


图 23.5　Tn3 两步转座简化图解。第一步，被 *tnpA* 基因产物催化，质粒（黑色）含有转座子（蓝色）与受体质粒（绿色，靶位点红色）融合形成共联体。在共联体形成时，转座子复制。第二步，被 *tnpR* 基因产物催化，使共联体分解为插入了转座子的受体质粒及原先含有转座子的供体质粒。

小结　许多转座子除含转座必需的基因外，还含有一些其他的基因，这些基因常常是抗生素抗性基因。如 Tn3 含有氨苄青霉素抗性基因。Tn3 及其相关的转座子的转座（复制式转座）有两个步骤。转座子首先复制，供体 DNA 整合到

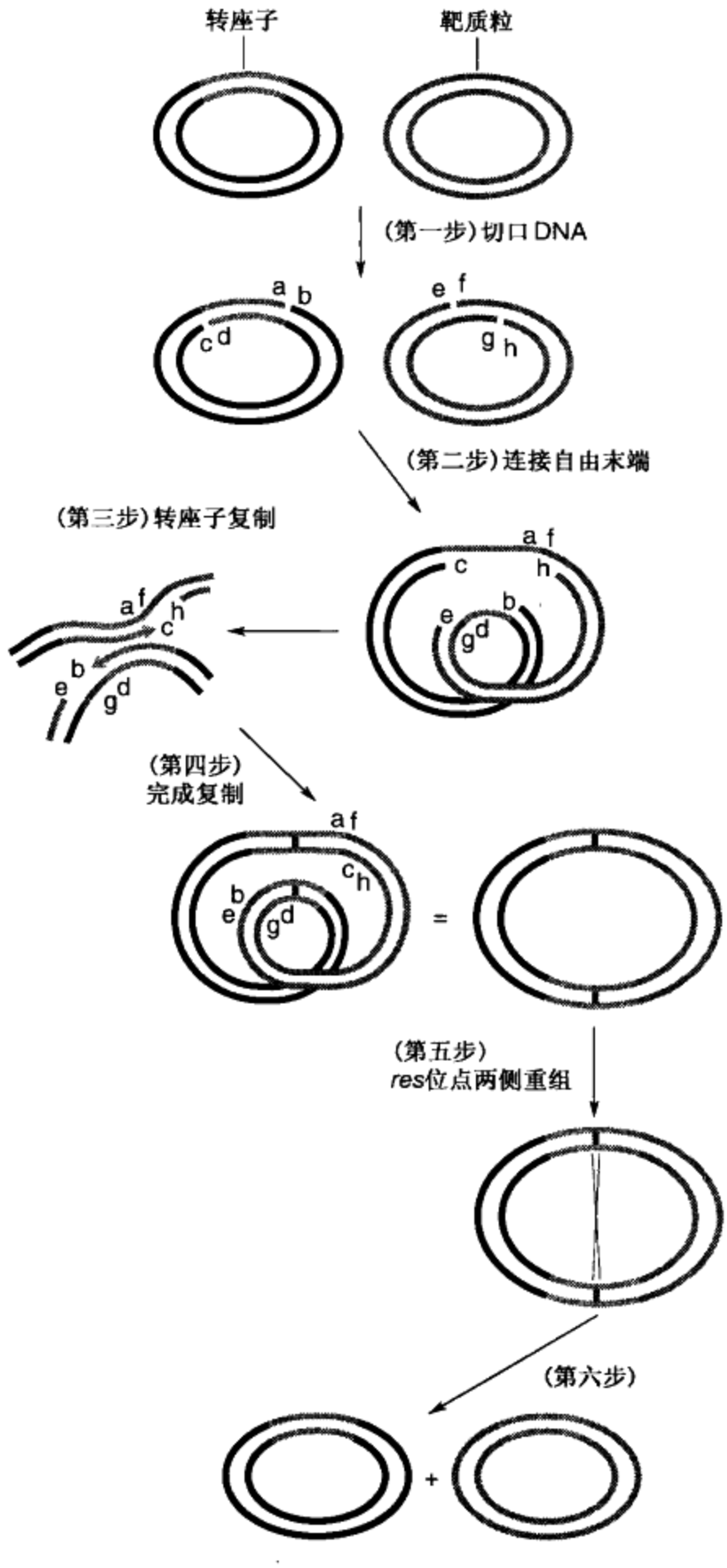


图 23.6　Tn3 转座的具体过程。第一步：两个质粒产生缺口形成自由末端，分别标记为 a～h。第二步：末端 a 和 f 相连接，g 和 d 相连接。b、c、e 和 h 仍是自由末端。第三步：保留的自由末端中的两个（b 和 c）作为 DNA 复制的引物，在复制区域的放大图中有显示。第四步：继续复制使末端 b 到达 e 而末端 c 到达 h。这些末端连接形成共联体。注意整个转座子（蓝色）被复制。这里第一次标识了成对的 *res* 位点（紫色），即使在前面的步骤中一个 *res* 位点已经存在。共联体中画了一个环，所以它的形成在先前的图中更清楚；但是，如果环是打开的，共联体将看上去像图 23.5 的样子。第五步和第六步：两个转座子拷贝的两个 *res* 位点间形成一个交叉，使得两个独立的质粒各自包含一个转座子的拷贝。以上过程显示了左右两个相同形式的共联体形成。

受体 DNA，形成共联体。然后，共联体被拆分成两个 DNA 环，每一个环各含一个拷贝的转座子。另一种不同于 Tn3 的途径是保守式转座（conservative transposition），转座子在转座过程中不发生复制。

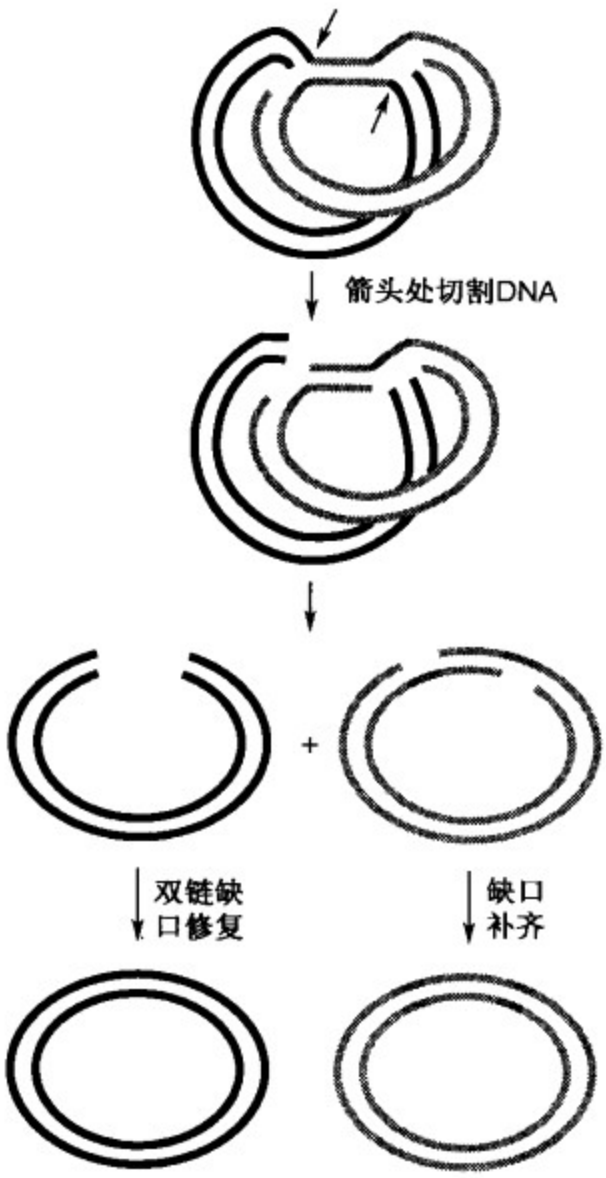


图 23.7　非复制式转座。开始的两步与复制式转座相同，顶部的结构与图 23.6 的第二、三步相同。接下来，在箭头处形成新的切开。这将供体质粒除了转座子以外的部分释放，而转座子整合到受体 DNA 上。间隙填补和切口封闭使靶质粒整合上它新的转座子。供体质粒的自由末端可以或不可以连接。无论哪种情况，供体质粒都失去了转座子。

23.2　真核生物的转座子

如果细菌是唯一含有转座因子的有机体，那将是不可思议的，因为这些转座因子会赋予其寄主强大的选择压。首先，许多转座子携带有对寄主有利的基因。因此，寄主能以消灭其竞争者为代价繁衍自身，并与寄主 DNA 的其他部分一起繁殖转座子。其次，即使转座子对其寄主不利，它们也可以在宿主中“自私”地复制。真核生物中确实存在转座因子。事实上，转座因子首先是在真核生物中被发现的。

第一例转座因子：玉米的 *Ds*/*Ac* 系统

20 世纪 40 年代后期，Barbara McClintock 在研究玉米时发现了第一个可以转座的因子。较长时间以来，人们已经在印第安玉米中观察到色斑籽粒的产生是由一种不稳定的突变所导致的。如图 23.8a 所示的有色籽粒。这种颜色是由玉米 *C* 位点编码的。图 23.8b 显示了当 *C* 基因发生突变时籽粒为白色而非紫色。图 23.8c 所示的斑点籽粒是在有些籽粒的细胞中发生了回复突变后的结果。无论何时发生回复突变，回复突变的细胞及其后代都能产生颜色，从而导致深色斑点籽粒的产生。很显然，籽粒中产生的许多斑点就意味着突变是很不稳定的：回复突变的频率远远高于我们对一般突变所预期的频率。

对这一现象的研究，McClintock 发现一个称为 *Ds*（解离因子）的转座因子插入到 *C* 基因中产生起始突变（图 23.9a 和图 23.9b）。另一个转座因子 *Ac*（激活因子）可以诱导 *Ds* 再次从 *C* 基因中转座出去，

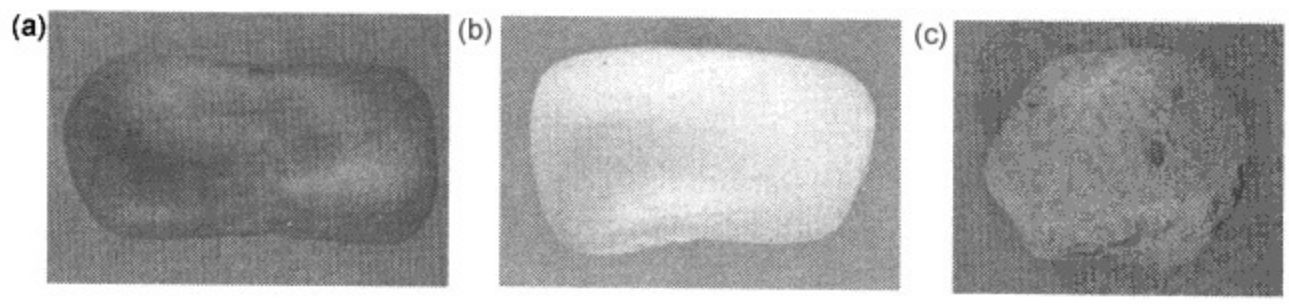


图 23.8　突变和回复突变对玉米籽粒颜色的影响。（a）野生型籽粒中有活性的 *C* 基因合成紫色素；（b）*C* 基因发生突变，阻碍色素的合成，产生无色籽粒；（c）部分细胞的 *C* 基因发生回复突变，重新激活色素合成，从而产生斑点籽粒。（*Source*: F. W. Goro, from Fedoroff, N., Transposable genetic elements in maize. *Scientific American* 86 (June 1984).）

产生回复突变（图 23.9c）。换言之，*Ds* 只有在 *Ac* 的帮助下才能发生转座。另一方面，*Ac* 是自主性转座因子，可以独自发生转座，不需要其他因子的帮助就能使靶基因失活。

在 McClintock 发现转座子数十年之后，我们可以利用分子生物学工具来分离和鉴定这些遗传因子。Nina Fedoroff 及同事获得了 *Ac* 和 3 种不同形式 *Ds* 的结构。*Ac* 类似于前文所介绍的细菌转座子（图 23.10）。*Ac* 长约 4500bp，含有一个转座酶基因和一段与转座酶的近末端重复区域邻接的、短的不完整的反向重复序列。*Ac* 缺失后形成不同形式的 *Ds*。*Ds-a* 和 *Ac* 十分相似，只是缺失了部分转座酶的基因。这一点可以解释 *Ds* 自身不能发生转座。*Ds-b* 的缺失片段更大，仅保留了转座酶基因的一个小片段，长度更短一些。*Ds-c* 则仅剩有 *Ac* 因子中反向重复序列和与转座酶结合的近末端重复区域。这些区域和片段都是 *Ds-c* 在 *Ac* 指导下转座所必需的靶位点。

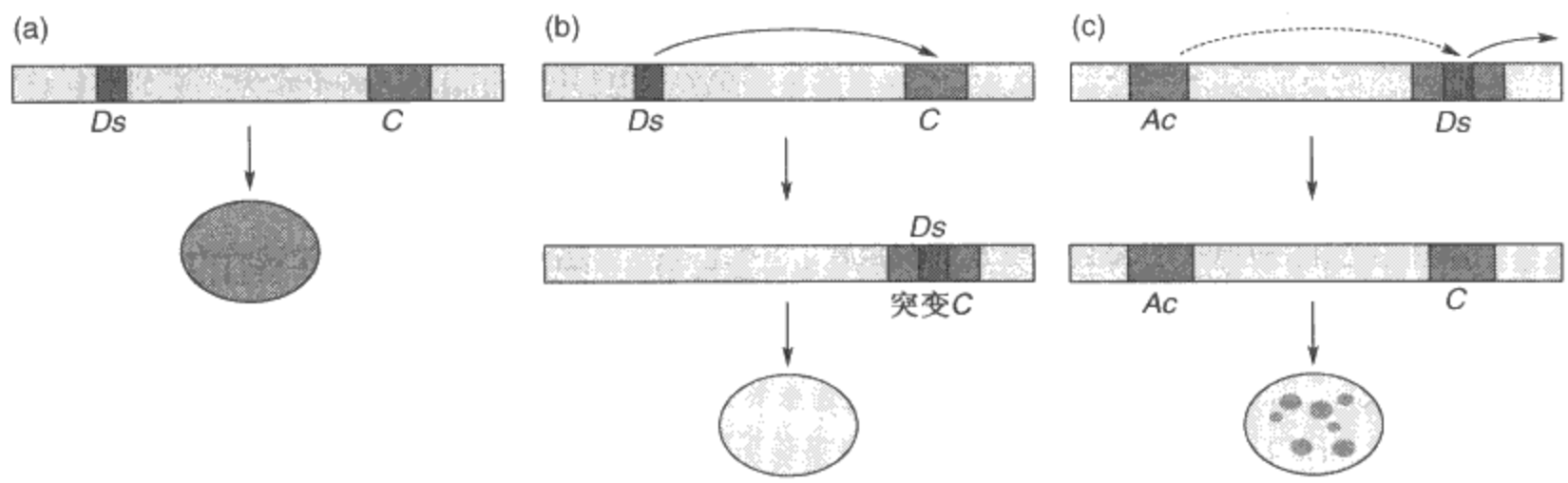


图 23.9　玉米中转座因子产生突变及回复突变。（a）野生型玉米籽粒包含有完整的活性 *C* 位点（蓝色），可以合成紫色素。（b）*Ds* 因子（红色）插入 *C*，使 *C* 失活，阻碍色素合成。因此籽粒呈无色。（c）*Ac*（绿色）和 *Ds* 因子同时存在使得 *Ds* 因子在许多细胞中可以跳出 *C* 位点，产生一些可以合成色素的细胞。这些细胞引起籽粒上出现紫色斑点。当然，*Ds* 必须在其缺损之前转座到 *C* 中，否则将需要 *Ac* 因子帮助。

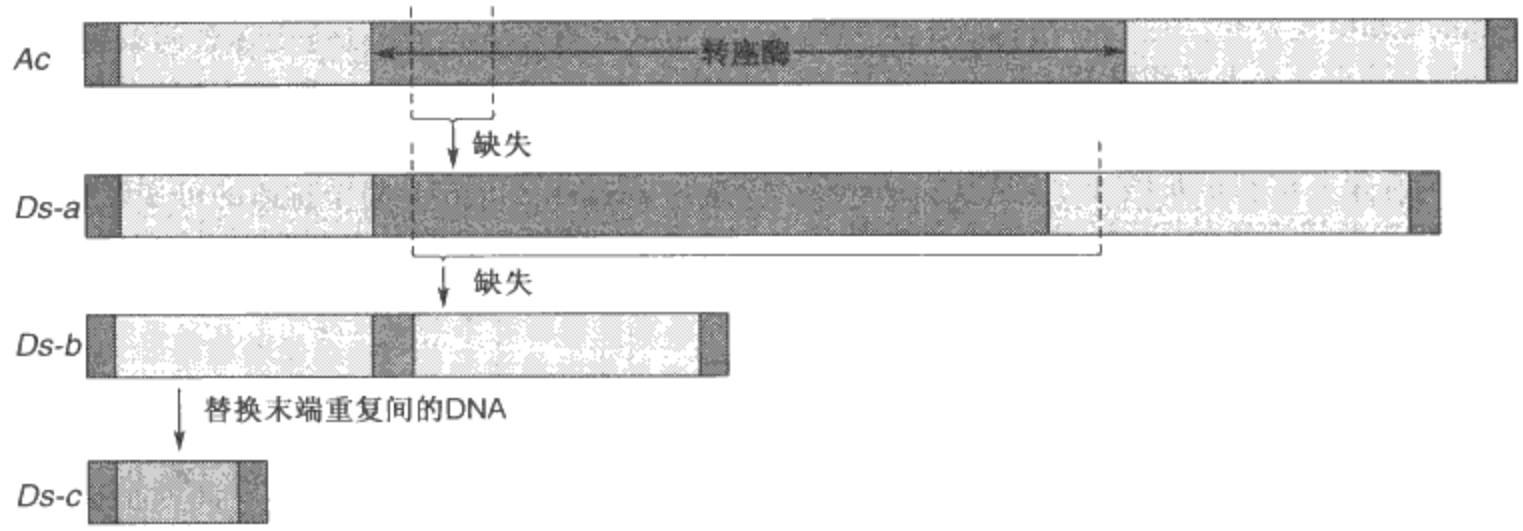


图 23.10　*Ac* 和 *Ds* 的结构。*Ac* 含有转座酶基因（紫色）和两个不完整的末端反向重复序列（蓝色），以及近末端重复区域。*Ds-a* 中缺失了转座酶基因中一段 194bp 的区域（虚线），其余部分与 *Ac* 相同。*Ds-b* 缺失了 *Ac* 中更大的片段。*Ds-c* 仅有末端反向重复序列和近末端重复区域与 *Ac* 相同。

有趣的是，孟德尔本人描述的第一个控制豌豆种子饱满和皱瘪的基因 *R* 或 *r* 似乎也含有转座因子。现在我们知道 *R* 基因编码一个参与淀粉代谢的酶（支链淀粉酶）。皱瘪表型是这个基因功能异常引起的，这种突变是由于在 *R* 基因中插入了一段类似于 *Ac*/*Ds* 家族成员的 800bp 的 DNA 片段引起的。

小结　玉米籽粒色斑是由 *C* 位点的非稳定突变的多次回复突变引起的，*C* 位点控制籽粒的颜色。*Ds* 因子转座到 *C* 基因中，使其突变，而后再从 *C* 基因中转座出去使 *C* 基因回复成野生型，从而导致了突变及回复突变的产生。*Ds* 自身不能发生转座，必须由自主因子 *Ac* 提供转座酶才能转座。*Ds* 因子或多或少缺失 *Ac* 因子中的一部分片段。所有的 *Ds* 要实现转座均需要一对末端反向重复序列和与其毗连的一段短的可以被 *Ac* 转座酶所识别的序列。

P 因子

杂种败育（hybrid dysgenesis）现象说明另外一种明显由真核生物转座子转座产生的突变。杂种败育是指在一个品系的果蝇与另一个品系的果蝇杂交产生的杂种后代中，染色体遭受多种损伤而使得杂种后代败育或不育。杂种败育是亲本双方共同作用的结果。例如，在 P-M 系统中，父本必须是 P 品系（父本贡献），母本必须是 M 品系（母本贡献）。M 父本与 P 母本的反交以及品系内的杂交（P×P 或 M×M）产生的子代都正常可育。

我们为什么怀疑此现象中涉及了转座子的作用呢？首先，任何 P 系雄性染色体都能导致与 M 系雌性杂交后代的败育。而且，来自部分 P 系雄性和部分 M 系雌性染色体的重组雄性染色体也常常引起后代的败育，表明这些染色体上的多个位点上都带有 P 性状。

对这种现象的一种可能的解释是 P 性状受到转座因子的控制，所以我们才会在许多不同的位点上都发现有 P 性状。事实上，这种解释是正确的。负责 P 性状的转座子称为 P 因子（P element），发现其仅存在于野生型果蝇中，在实验室培养的品系中除了生物学家专门导入的以外，均没有发现该因子。Margaret Kidwell 及同事研究发现，P 因子插入到了败育果蝇的 *white* 位点中。他们发现这些因子的碱基序列高度一致但大小上区别很大（500～2900bp）。此外，P 因子具有末端反向重复，两侧有短的寄主 DNA 的正向重复，这些都是转座子的特征。最后，*white* 突变体在丢掉完整的 P 因子后表现出很高的回复突变频率，这是转座子的又一特性。

如果 P 因子与转座子相类似，那为什么它们仅在杂种中才转座并引起败育呢？答案就是 P 因子还编码一个转座抑制物，在发育的生殖细胞的细胞质中累积。因此，P 或 M 的雄性与 P 的雌性杂交，雌性细胞质中含有抑制物，能与 P 因子结合并阻碍其发生转座。但是在 P 的雄性与 M 的雌性的杂交中，早期胚胎中不含该抑制物，而由于 P 因子仅在发育中的生殖细胞中有活性，所以初始阶段也并不形成抑制物。最后当 P 因子被激活时，转座酶和抑制物同时产生，但是仅有转座酶进入细胞核，在细胞核中它可以激发转座的发生。

杂种败育在物种形成中有着重要作用——产生一些不能相互杂交的新物种。同一物种的两个品系（如 P 和 M）经常产生不育的后代倾向于形成生殖隔离——与通常的情况不同，它们的基因之间不再混合，最终由于它们在遗传上的差异导致完全不能进行杂交繁育，进而成为两个分离的物种。

目前，P 因子通常被用作果蝇遗传实验中的诱变剂。这种方法的一个优点是突变容易被定位，只要寻找到 P 因子也就找到了被阻断的基因。分子生物学家也将 P 因子用于果蝇的转化，即通过 P 因子将改造过的基因转化到果蝇中。

小结 果蝇杂种败育的 P-M 系统是由两个相连的因子引起的：①父本提供的转座因子 P；②母本提供的 M 细胞质，这种细胞质能使 P 因子发生转座。P 系雄性和 M 系雌性的杂交后代中 P 因子多次发生转座。转座引起染色体的损伤、突变，导致杂种后代的不育。P 因子作为诱变剂和转化因子在果蝇的遗传实验中有一定的实用价值。

23.3 免疫球蛋白基因的重排

哺乳动物 B 细胞的基因重排产生抗体（antibody）及免疫球蛋白（immunoglobulin）的过程和 T 细胞的基因重排产生 T 细胞受体（T-cell receptor）的过程均与转座过程类似，甚至两个过程中的重组酶也有相似的结构。因为有这么多相似点，我们将重排也纳入本章的内容。

在第 3 章中曾提到抗体由 4 条多肽链组成：两条重链和两条轻链（T 细胞受体含两条重 β 链和两条轻 α 链）。图 23.11 中示意了抗体的结构，并标明了与入侵抗原的结合位点。这些位点我们称为**可变区**（variable region），每个抗体的可变区都不一样，构成了抗体的特异性；蛋白质的其余部分（保守区，

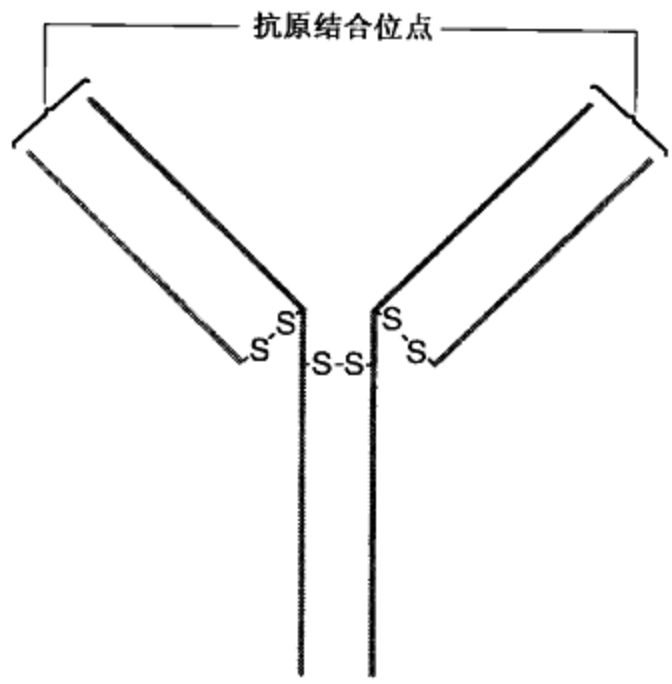


图 23.11 抗体的结构。抗体由两条轻链（蓝色）通过二硫键连接两条重链（红色）组成，两条重链之间也是由二硫键连接的。抗原结合位点位于蛋白质链可变区的末端氨基酸处。

constant region）尽管在少数不同种类的抗体间有所不同，但在同类抗体中该区是不变的。每个免疫细胞都只能产生一种特定的抗体。更为明显的是，人类免疫细胞能产生不同的抗体以抵御任何外来物质的入侵，意味着能产生上百万种不同的抗体。

那么这是否意味着有上百万种不同的抗体基因呢？这种假设是站不住脚的：它将造成我们的基因组为了携带所有这些必需的基因而负担沉重。那么如何解决抗体多样性的问题呢？不像看上去的那样，成熟的 B 细胞作为专门产生抗体的细胞，通过基因组重排将分散的抗体基因连接起来。将异源区段随机选择的基因连接起来的方法就像是从午宴菜单上点菜一样（从 A 栏选择一项，再从 B 栏选择一项）。重排大大增加了基因的可变性。例如，如果 A 栏有 41 个选项，B 栏有 5 个选项，A＋B 栏的组合总数为 41×5，即 205。因此，46 个基因片段可以组合成 205 个基因，这仅仅是对一条抗体多肽链而言。如果其他链也用同样的方法，抗体的总数将是两种多肽链数的乘积。以上说法尽管在理论上是成立的，但还是将抗体基因的实际情况大大简化了；正如我们即将看到的，关于抗体多样性，有更复杂的机制可大大增加抗体产物的数量。

对哺乳动物抗体的研究表明抗体轻链的两个家族分别称为 Kappa（κ）和 lambda（λ）。图 23.12 显示了人类部分 κ 轻链的基因重排。“菜单栏”的“A 栏”包含 41 个可变区部分（V）；“菜单栏”的“B 栏”包含 5 个连接区部分（J）。J 片段实际上编码可变区的最后 12 个氨基酸，但是它们距离 V 区段的其余部分较远，而离一个保守区较近。在抗体形成细胞进行分化之前，以及将不相连的区域连在一起的重排行为发生之前，生殖细胞中的 J 片段一直维持着上述状态。图 23.12 描绘了重排和表达事件。

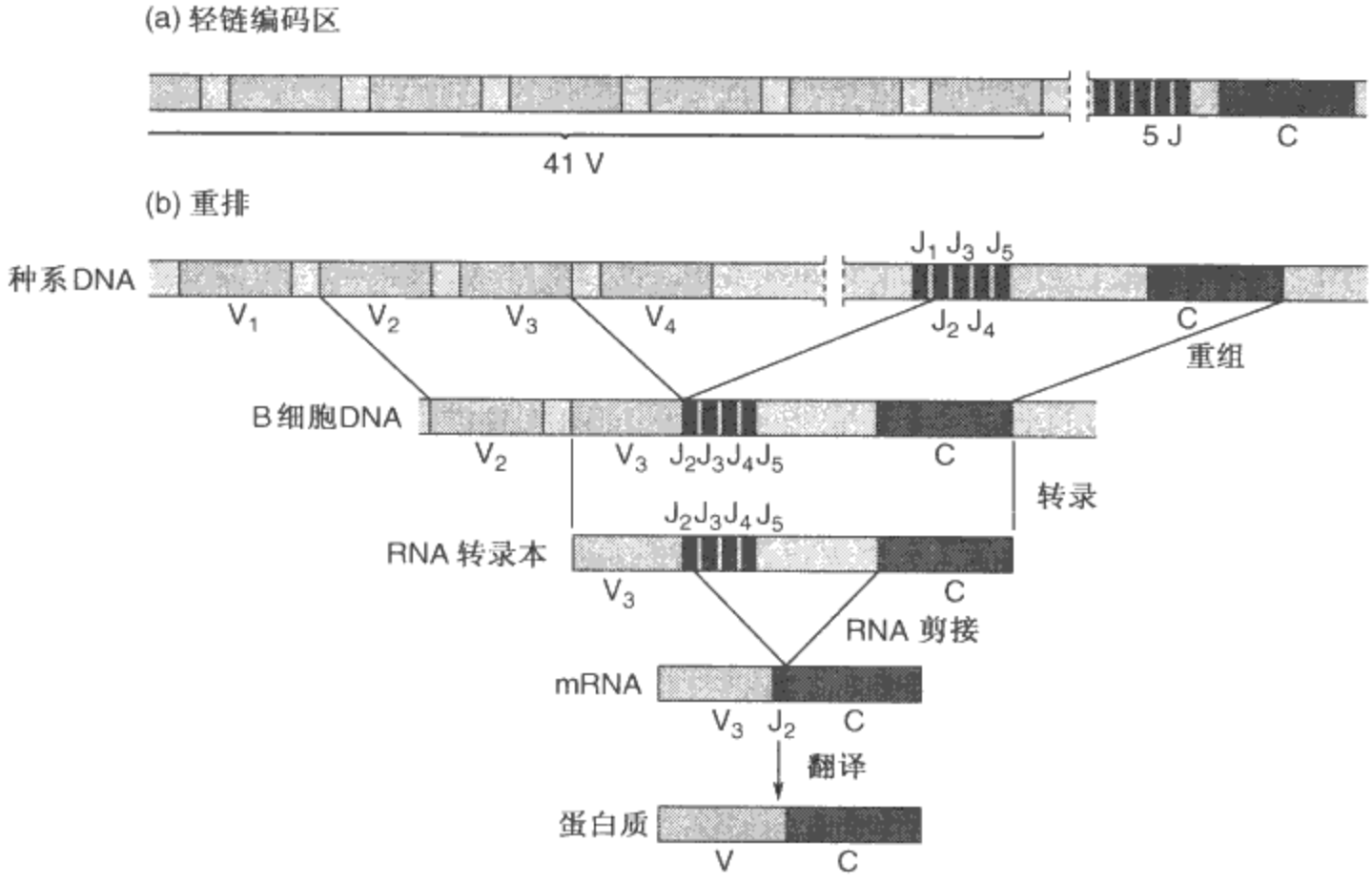


图 23.12　一个抗体轻链基因的重排。（a）人类 κ 抗体轻链编码 41 个不同的基因片段（V：浅绿色）、5 个连接片段（J：红色）和一个保守片段（C：蓝色）。（b）在抗体形成细胞的成熟过程中，一段 DNA 缺失，将 V 片段（此例中的 V_3）与 J 片段（此例中的 J_2）相连接。这个基因可以转录为这里显示的包括 J 片段和插入序列的前体 mRNA。将 J_2 与 C 之间的片段剪切后成为成熟的mRNA，翻译成如底部所示的抗体蛋白。mRNA 中的 J 片段翻译成抗体可变区的一部分。

首先，一段 V 区和一段 J 区发生重组。在此情形下，虽然看起来 V_1 和 J_4 连接更容易，但还是 V_3 和 J_2 相连接，选择是随机的。基因的两个部分组装后，从 V_3 的起点开始转录到 C 的末端。其次，通过剪接将 J_2 区和 C 转录物连接，移除 J 区的额外部分和位于 J 区和 C 之间的间隔序列。重排发生在 DNA 水平，而发生在 RNA 水平上的剪切机制已在第 14 章介绍过。组装好的 mRNA 转移到细胞质中，翻译成抗体的轻链，轻链上有可变区（由 V 和 J 共同编码）和保守区（由 C 编码）。

为什么转录从 V_3 区段起点开始而不是在更上游的地方呢？答案似乎是位于 J、C 区之间的内含子

中的一个增强子激活了与其最邻近的一个启动子，此例中为 V_3 启动子。这也便于激活重排之后的基因，只要增强子与启动子的距离近到足以开启启动子即可。

重链上基因的重排更加复杂，因为在 V 区和 J 区之间还有一套额外的基因。这些基因片段被称为 D 区，即多样性（diversity），在我们的“菜单栏”上表示为第三栏。图 23.13 显示了重链 48 个 V 区、23 个 D 区和 6 个 J 区的组装。以此为基础，一个细胞可以组成 48×23×6，即 6624 个不同的种类基因。进而 6624 个不同的重链与 205 个 κ 轻链及 170 个 λ 轻链相结合，可以形成 2.5×10^6 种抗体或可变区的结合体。

抗体多样性的形成还有其他的来源。首先，

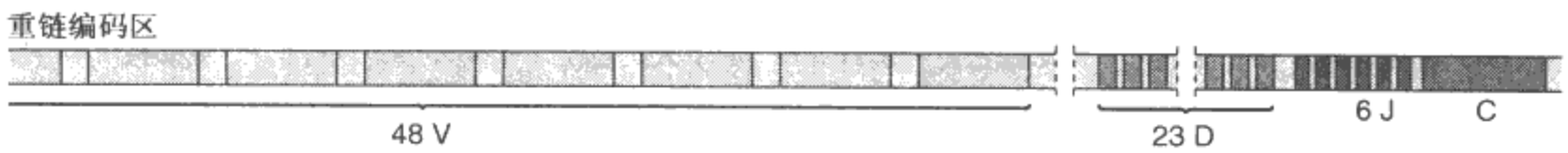


图 23.13　抗体重链编码区的结构。人类重链编码 48 个可变区（V：浅绿色）、23 个多样区（D：紫色）、6 个连接区（J：红色）和 1 个保守区（C：蓝色）。

V（D）J连接［V（D）J joining］的各区段连接机制并不十分精确，可导致在连接位点的两侧增加或减少若干碱基，从而产生抗体氨基酸序列的差异。

抗体多样性的另一个来源是体细胞超突变或体细胞（非性细胞）的快速突变。在此情形下，抗体基因产生突变的同时形成抗体的 B 细胞也进行繁殖以应对入侵者的挑战。

遗传学和生物化学分析表明体细胞超突变的发生分两步进行。第一步，DNA 复制时，B 细胞活化、诱导产生的胞苷脱氨酶使胞嘧啶脱氨基生成尿嘧啶。第二步，尿嘧啶进入错配修复过程引起突变或者经由尿嘧啶-*N*-糖苷酶转移尿嘧啶，留下无碱基位点。无碱基位点通过 DNA 聚合酶 ζ、η、θ（可能还有 ι）参与的跨越损伤支路进行修复（第 20 章）。这些聚合酶易于产生倾向错误的修复，尤其是在无碱基位点，因此能产生很多突变。

总之，基因片段的不精确连接和体细胞超突变极大地扩大了抗体的数量。实际上，生物体一生的抗体总数估计高达 1000 亿。这应该足以对付各种进攻者。

小结　脊椎动物的免疫系统能产生 10 亿种不同的抗体与外来物质发生作用。这些免疫系统通过 3 个基本机制产生了巨大的多样性：①2 个或 3 个组成部分被用来组装抗体基因的轻链和重链，每个部分都分别从自身的异质源库中选择；②通过减少甚至增加额外碱基对的非精确连接机制来重组基因的各个部分从而改变基因；③可能在免疫细胞的增殖过程中引起体细胞的高频突变，产生差异轻微的基因。

重组信号

重组体系是怎样决定在什么位置对免疫球蛋白基因的不同部分进行剪接的呢？测定了许多鼠免疫球蛋白基因（编码 κ、λ 轻链和重链）的序列后发现了一种共有的模式（图 23.14a）：相邻的两个编码区之间有一个保守的七聚体回文序列 5′-CACAGTG-3′。这个七聚体还伴随着一个 5′-ACAAAAACC-3′的九聚体保守序列。由一个 12 个碱基对（12 信号，12 signal）或 23（±1）碱基对（23 信号，23 signal）的非保守区将七聚体和九聚体分隔开。这些重组信号序列（recombination signal sequence，RSS，图 23.14b）的重排常常是将一个 12 信号连接到一个 23 信号上。这种 12/23 法则（12/23 rule）保证了 12 信号和 23 信号自身不会连接在一起，以确保只有一个编码区可以被整合到成熟的免疫球蛋白基因中。

除了 RSS 保守序列的存在外，还有没有其他的证据表明它的重要性呢？Martin Gellert 等采用碱基替代法系统地造成七聚体和九聚体发生突变，增加或减少间隔区的碱基，然后观察这些改变对重组的影响。用以下方法测定重组的有效性：构建一个如图 23.15 所示的重组质粒。第一个元件是 *lac* 启动子，后面依次接一个 12 信号、原核生物转录终止子（Transcription terminator）、一个 23 信号，最后连接一个 *cat* 报告基因。制备覆盖整个 RSS 的突变体库，然后将改造后的质粒引入 pre-B 细胞系。最后，将从 pre-B 细胞中分离纯化的质粒导入氯霉素敏感性 *E. coli* 细胞中，检测氯霉素抗性。如果没有发生重组，转录终止子将阻碍 *cat* 基因的表达，几乎没有氯霉素的抗性。若重组发生在 12 信号与 23 信号之间，终止子将会因被倒位或缺失而失活，在此情况下，*lac* 启动子将启动 *cat* 基因的表达，形成许多抗氯霉素的克隆。该实验表明七聚体或九聚体的大量碱基突变将使重组的有效性降低到本底水平。在间隔区插入或缺失碱基的结果与之相同。所以，RSS 的每个元件对 V（D）J 的重组都很重要。

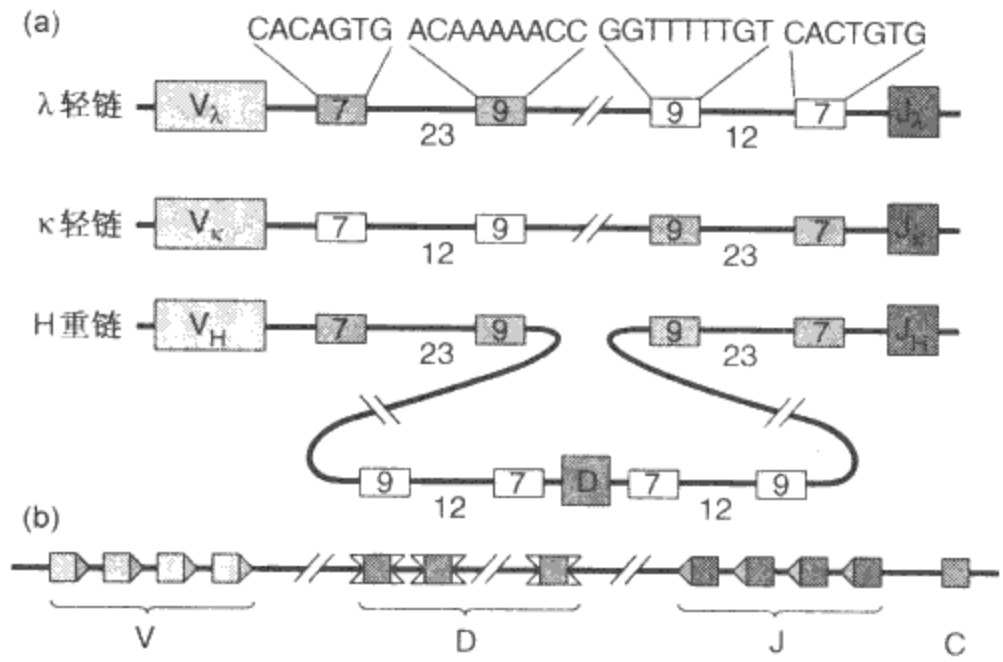


图 23.14　V（D）J 连接信号。（a）免疫球蛋白的 κ、λ 轻链基因和重链基因的编码区周围的信号重排。标记“7”和“9”的框分别是保守的七聚体和九聚体。方框上面给出了它们各自的保守序列。12 和 23 个碱基对间隔区已被标记。注意 12 信号和 23 信号两两相连发生重排后自然组装成一个完整的基因。（b）免疫球蛋白重链基因的 12 和 23 信号重排示意图。黄色符号代表 12 信号，橙色三角形代表 23 信号。再次注意 12/23 法则是如何确保在重排的基因中只包有每个编码区（V、D、J）中的一个。（*Source*：（a）Adapted from Tonegawa，S.，Somatic generation of antibody diversity. *Nature* 302：577，1983.）

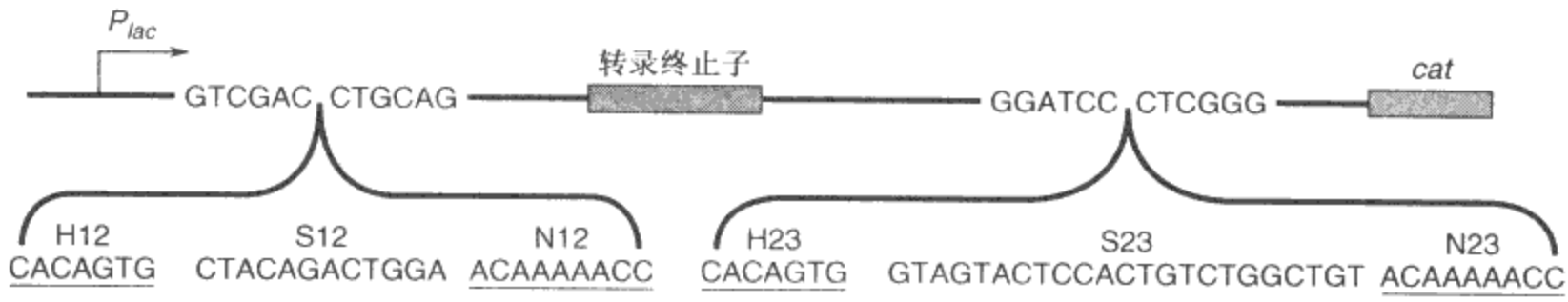


图 23.15　测定 RSS 突变体对重组有效性影响的报告基因载体的结构。Gellert 等构建的重组报告质粒包含一个 *lac* 启动子和 *cat* 基因，两者之间插入一个两侧有一个 12 信号和一个 23 信号的转录终止子。两个 RSS 之间发生重组后将会引起终止子倒位或缺失，由此启动 *cat* 基因的表达。一方面，重组质粒转化细菌细胞后产生大量的 CAT 产物克隆，这些克隆表现出抗氯霉素特性；另一方面，若是非重组质粒转化的细菌将不产生抗氯霉素的克隆。（*Source*：Adapted from Hesse，J.，M. R. Lieber，K. Mizuuchi and M. Gellert，V（D）J recombination：a functional definition of the joining signals. *Genes and Development* 3：1053-61，1989.）

小结　V（D）J 重组体的重组信号序列（RSS）由 12bp 或 23bp 隔开的一个七聚体和一个九聚体组成。重组只能发生在 12 信号和 23 信号之间以确保重组基因中仅整合每个编码区中的一个。

重组酶

David Baltimore 及同事研究编码 V（D）J 重组酶的基因时所用质粒类似于以上已介绍的重组报告质粒，但通过让其获得霉酚酸药物抗性且使其在真核细胞中便于操作。将这种质粒连同鼠基因组 DNA 片段转入 NIH 3T3 细胞，这种细胞缺失 V（D）J 重组活性，通过分析 3T3 细胞的药物抗性来检测重组。这样就可以鉴定出在体内激发 V（D）J 连接活性的重组活性基因（*RAG-1*）。

但是，用包含几乎全部 *RAG-1* 的基因组克隆的重组激发水平为中等，和全基因组 DNA 的重组激发水平几乎一致。而且，用包含全部 *RAG-1* 序列的 cDNA 克隆效果也不好，所以可能漏掉了一些元件。Baltimore 研究小组对包含几乎全部 *RAG-1* 基因的基因组片段测序，发现了一个与之紧密相连的基因。为了回答这个基因是否与 V（D）J 连接有关，他们在相同的转染实验检测中加入 *RAG-1* cDNA 的基因组片段。当两个 DNA 同时导入一个细胞时，得到了药物抗性的细胞。由此，他们发现了两个与 V（D）J 重组相关的基因，并把第二个命名为 *RAG-2*。

RAG-1 和 *RAG-2* 仅在 pre-B 和 pre-T 细胞中表达，这两种细胞分别是 V（D）J 连接成免疫球蛋白和生成 T 细胞受体基因的场所。T 细胞受体与免疫球蛋白结构相似，是与膜结合的抗原结合蛋白。编码 T 细胞受体的基因和免疫球蛋白的基因遵照相同的重排法则，即由包含 12 信号和 23 信号的 RSS 完成。因此，*RAG-1* 和 *RAG-2* 显然与免疫球蛋白和 T 细胞受体的 V（D）J 连接有关。

V(D)J 重组机制

V（D）J 的非准确连接导致了连接过程中产物的多样性。在连接处常常发生碱基的缺失和额外碱基的增加，这一点对免疫球蛋白和 T 细胞受体的形

成是有利的，因为这样可以用有限的基因组合获得类型更多的蛋白质。

如何解释这种连接的非准确性呢？图 23.16 显示了在两个编码片段间的 RSS 两侧的插入片段被切除的机制。*RAG-1* 和 *RAG-2* 的基因产物分别是 RAG-1 和 RAG-2，在 DNA 连接处先切出一个切口。然后新的 3′-羟基进攻互补链上的磷酸二酯键，释放出插入片段并在编码片段末端形成发夹结构。这些发夹结构是非精确连接的关键，可以开启发夹结构任意一边的顶端而增、减碱基，使 DNA 形成连接平末端。Rag-1 和 Rag-2 蛋白把两个发夹结合成为一个复合体，相互间进行共价连接。

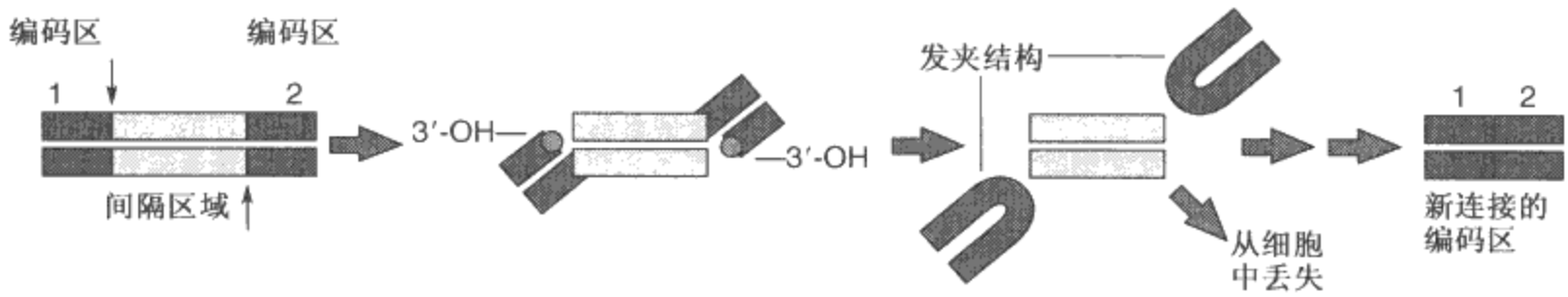


图 23.16　RSS 的剪切机制。在连接编码区（红色）和插入片段（黄色）的 RSS 的两条链的对应位置产生切口（垂直箭头）。新的 3′-羟基（蓝色）攻击并打断互补链形成发夹结构并释放出插入片段，插入片段丢失。最后发夹结构被打开，两个编码区用非精确机制连接起来。（*Source*：Adapted from Craig, N. L., V (D) J recombination and transposition: closer than expected. *Science* 271: 1512, 1996.）

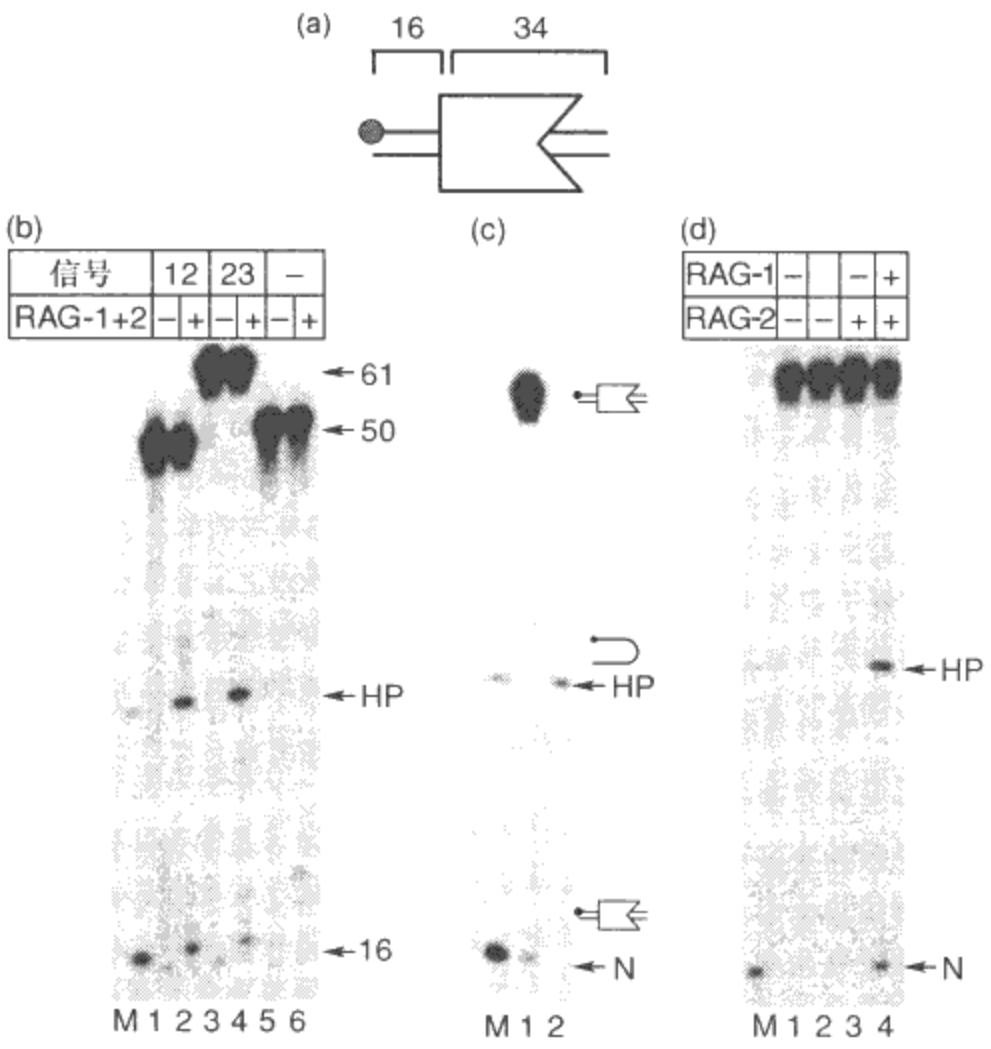


图 23.17　剪切产物的鉴定。（a）剪切底物。Gellert 及同事构建了这个 50 聚体的标记物，包括左侧一段 16bp DNA、一个 12 信号（黄色）和右侧一段包含 12 信号的 34bp 片段。单链 5′端标记用红点表示。研究者还构建了一个类似的含有一个 23 信号的长为 61 聚体的底物。（b）鉴定发夹结构产物。如上部方格所示，将 RAG-1 蛋白和 RAG-2 蛋白与标记 12 信号或 23 信号的底物温育。温育后产物用非变性凝胶电泳并以放射自显影检测标记产物。右侧标出了 61 聚体、50 聚体底物，发夹结构（HP）和 16 聚体的显影位置。（c）非变性凝胶上的产物检测。Gellert 研究小组从非变性凝胶板上回收标记产物（显然没有切割的 50 聚体和 16 聚体片段）。回收产物分别点在变性凝胶泳道 1 和泳道 2 电泳，另一个泳道点与非剪切底物，16bp 发夹结构（HP）和从变性的底物缺口释放出的单链 16 聚体对应的标记（Marker）。（d）对 RAG-1 和 RAG-2 的需求。这个实验与 b 相似，唯一的区别是只有 RAG-1 蛋白和 RAG-2 蛋白这两个变量（如上部方格所示）。“N”表示从切割底物中释放出的 16 聚体的位置。（*Source*：McBlane, J. F., D. C. Van Gent, D. A. Ramsden, C. Romeo, C. A. Cuomo, M. Gellert and M. A. Oettinger, Cleavage at a V (D) J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. *Cell* 83 (3 Nov 1995) f. 4 a-c, p. 390. Reprinted by permission of Elsevier Science.）

如何知道形成了发夹结构呢？首先在活体内发现了浓度极低的发夹结构。随后，Gellert 小组发展出一种便于观察的体外系统。图 23.17a 为该研究小组制备的一个标记底物，长 50 个核苷酸，^{32}P 标记 5′端，包含一个 12 信号（黄色标记），左侧有一段 16 碱基对的片段，右侧末端有一段包含 12 信号的 34 碱基对的片段。制备的另一个相似的底物包含同样的侧翼序列，但含有 23 信号而不是 12 信号，所以它的长度为 61 碱基对。

研究小组将这些底物分别与 RAG-1 和 RAG-2 及与人类同源的鼠的 RAG-1 和 RAG-2 一起温育，然后在非变性条件下电泳检测产物是否发生了 DNA 剪切（图 23.17b）。结果发现一个 16 聚体的片段，表明双链发生了剪切。但是非变性凝胶电泳无法区分一段 16 聚体的双链和一个带有发夹末端的 16 聚体片段，所以研究者在较高温度下在含有尿素的变性聚丙烯酰胺凝胶中电泳检测产物（图 23.17c）。在这样的条件下，16 聚体的双链 DNA 将成为两条 16 聚体的单链 DNA，若是带有发夹结构末端的 16 聚体片段则会成为一条 32 聚体的单链 DNA。在 RAG-1 蛋白和 RAG-2 蛋白都存在的情况下不论 DNA 含有的是 12 信号还是 23 信号，研究小组观察到的都是后者。若 DNA 中没有 12 信号或 23 信号则没有产物和发夹结构，或者反应体系中缺少 RAG-1 蛋白或 RAG-2 蛋白也没有产物（图 23.17d）。因此，RAG-1 和 RAG-2 对 12 信号和 23 信号均能识别，在毗连信号处剪切 DNA，在编码片段末端形成发夹结构。

然而，在非变性凝胶上的 16 聚体产物，在变性凝胶上仅是发夹结构产物，说明没有 16 聚体的双链形式。但在非变性凝胶上和底物一起移动的标记的 DNA 在变性凝胶上会产生双链 16 聚体，这不可能由双链断裂而来，否则不会在非变性凝胶上与底物一起保留下来。因此，它一定是来源于标记链上的切口。由切口产生的 16 聚体在非变性凝胶电泳过程中保留与其互补链碱基配对，但在变性凝胶电泳中将和 16 聚体一样独立泳动。因此，单链切口显然也是 RAG-1 蛋白和 RAG-2 蛋白作用的一部分。

为了深入研究切口和发夹结构形成之间的关系，Gellert 等进行了时序动态的研究，设置底物与 RAG-1 蛋白和 RAG-2 蛋白的不同温育时间，然后变性凝胶电泳检测产物。他们发现首先出现切口型，接着是发夹型。这表明切口型是发夹型的一个前体。为了验证这种假设，他们构建了一个切口中间体并与 RAG-1 和 RAG-2 一起温育。有足够的证据表明，RAG-1 和 RAG-2 将切口 DNA 转化成发夹结构。Gellert 研究小组随后的工作表明事件发生的顺序可能是这样的：RAG-1 和 RAG-2 在一条 DNA 链毗连 12 信号或 23 信号的地方切出一个切口；然后新形成的羟基末端在酯基转移反应中攻击另一条链，形成如图 23.16 所示的发夹结构。

是什么酶打开了由 RAG-1 和 RAG-2 产生的发夹结构的呢？Michael 等 2002 年证实是一种称为 Artemis 的酶在起作用。Artemis 自身有核酸外切酶活性。但是与 DNA-PK$_{cs}$结合后，Artemis 将获得内切核酸酶活性从而切开发夹结构。在第 20 章中我们讨论修复双链 DNA 断裂过程中非同源 DNA 末端连接（NHEJ）时就介绍过 DNA-PK$_{cs}$。事实上，打开的发夹结构的连接与 NHEJ 很相似，均依靠 NHEJ 机制。

与免疫球蛋白基因重排类似，T 细胞受体重排过程中产生的发夹结构也要用 Artemis 切开。没有抗体，B 细胞就没有作用，同样的，没有 T 细胞受体，T 细胞也是无用的。因此，缺失 Artemis 的功能就意味着同时丧失了 B 细胞和 T 细胞的功能。事实上，Artemis 基因缺陷的人具有严重的联合免疫缺陷（SCID，“气泡男孩”综合征），并且对任何病原物都不会产生免疫反应。他们必须与世隔绝才能存活下来。

小结 RAG-1 和 RAG-2 在邻近 12 信号或 23 信号的 DNA 单链上切开一个切口，导致转酯反应发生，产生的3′-羟基攻击互补链并使其断裂，在编码片段的末端形成一个发夹结构。然后发夹结构以非精确方式断裂，通过碱基的增、减连接编码区。

23.4 反转录转座子

McClintock 发现的玉米转座子类似于本章前面介绍的细菌转座子，是剪贴式或复制粘贴式转座子中的一种。如果 DNA 复制也包括在内，那么它是直接复制的。人类也带有这类转座子，它占人类基因组的 1.6%。最常见的是 mariner，但是至今研究过的所有 mariner 因子在转座过程中都是有缺陷的。真核生物还携带有另外一种转座子——反转录转座子（retrotransposon），它通过 RNA 中间体进行复制。因此，反转录转座子类似于反转录病毒，有些反转录病毒引发脊椎动物产生肿瘤，有些反转录病毒（人类免疫缺乏病毒或 HIV）引发 AIDS。在介绍反转录转座子的复制机制之前，我们先了解一下反转录病毒的复制。

反转录病毒

反转录病毒最显著的特点是能以其 RNA 基因组复制 DNA，并由此而得名。从 RNA 到 DNA 的反应是转录过程的一个逆反应，通常称为反转录

(reverse transcription)，Howard Temin 和 David Baltimore 同时在 1970 年向持有疑问的科学界证明确实可以发生这种逆反应。他们通过在病毒颗粒中找到了能催化反转录反应的酶证明了这一点。因此，这种酶就被命名为反转录酶（reverse transcriptase）。更准确的名称应为依赖 RNA 的 DNA 聚合酶（RNA-dependent DNA polymerase）。

图 23.18 阐述了反转录病毒的复制周期。开始是一个病毒感染细胞。病毒含有 RNA 基因组的两个拷贝，在其 5′端通过碱基互补配对连接（为了简便，此处只显示了一个拷贝）。当病毒进入细胞时通过反转录酶的作用将末端带有长末端重复序列（long terminal repeat，LTR）的病毒 RNA 反转录成双链 DNA。这段 DNA 与寄主基因组融合，形成病毒基因组的整合形式，称为原病毒（provirus）。寄主 RNA 聚合酶 II 转录原病毒产生病毒 mRNA 进而翻译成病毒蛋白。在完整的复制周期中，聚合酶 II 也产生原病毒的 RNA 拷贝形成新的病毒基因组。基因组 RNA 包装成病毒颗粒（图 23.19）后出芽脱离被感染的细胞，再去感染其他细胞。

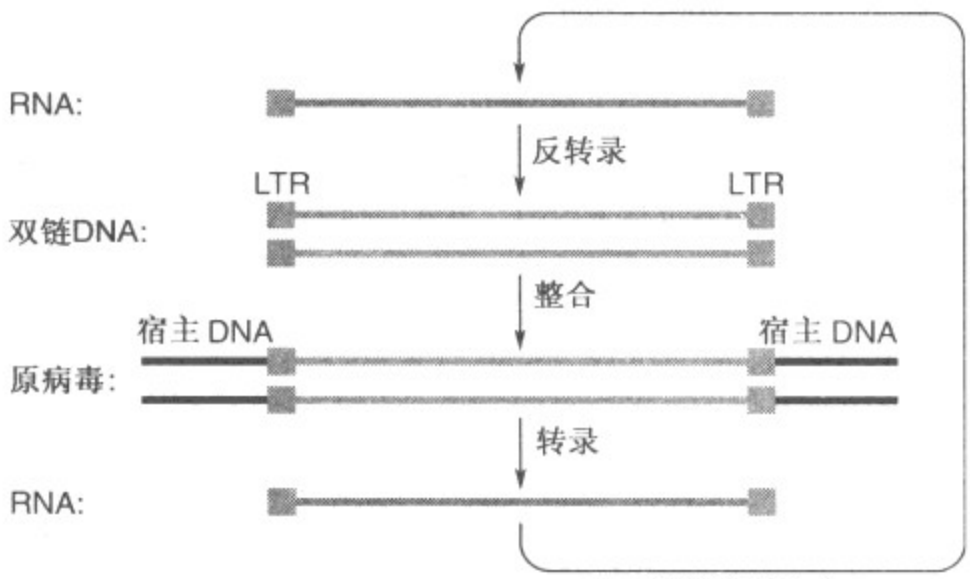


图 23.18　反转录病毒的复制周期。反转录病毒的基因组是 RNA，在末端携带有一段长的末端重复序列（LTR；绿色）。反转录酶催化 RNA 复制出线状双链 DNA 拷贝，双链 DNA 将整合到寄主 DNA（黑色）上，成为原病毒形式。寄主的 RNA 聚合酶 II 转录原病毒形成基因组 RNA。病毒 RNA 包装成病毒颗粒（packaging into virus），从细胞中出芽脱离（budding），然后感染另一个细胞，开始新一轮周期。

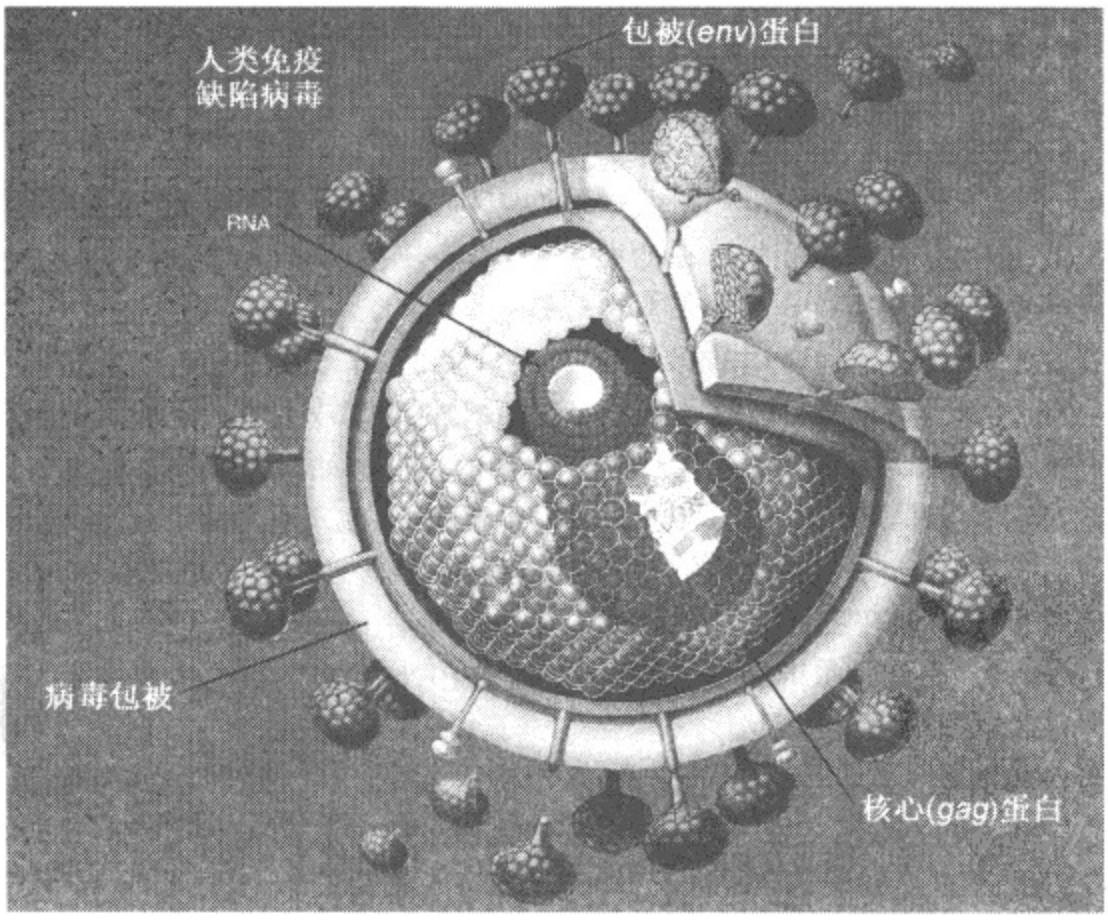


图 23.19　一种反转录病毒——人类免疫缺陷病毒（HIV）模型。病毒中央是 RNA 基因组（紫色）。外面包被一层由 *gag* 基因产物——核心蛋白质（core protein）（灰色）组成的外壳。病毒中心外包裹着细胞质膜产生的包被（viral envolope）（橙色）。包被上结合有病毒 *env* 基因的产物——包被蛋白（envolope protein）（绿色）。在 RNA 肿瘤病毒中也发现了同源基因，RNA 肿瘤病毒颗粒也可以用相同的模型表示。（*Source*：Courtesy Coulter Corporation.）

反转录酶的证据 对于反转录现象的质疑起源于从来没有被人观察到的这种反应，这也违背了Watson和Crick提出的分子生物学的“中心法则”，中心法则提出遗传信息是从DNA流到RNA再到蛋白质的，不能逆向流动。后来Crick说明DNA→RNA的箭头应该是双向的，但在当时显然并不能被学术界理解和接受。Baltimore和Temin掌握了什么证据足以消除这些疑虑呢?

图23.20显示了Baltimore的实验结果。将纯化的反转录病毒颗粒（Raucher鼠白血病病毒，R-MLV）与4种dNTP一起温育，其中包括[^{3}H] dTTP，然后通过酸沉淀法测定结合到多聚体（DNA）中的[^{3}H] dTTP标记。他观察到明显的[^{3}H] dTTP的结合（红色曲线）可以被反应体系中加入的RNase抑制（蓝色曲线），和RNase进行预温育会产生更明显的抑制（绿色曲线）。对RNase的敏感性与RNA是反转录反应模板的假设相一致。

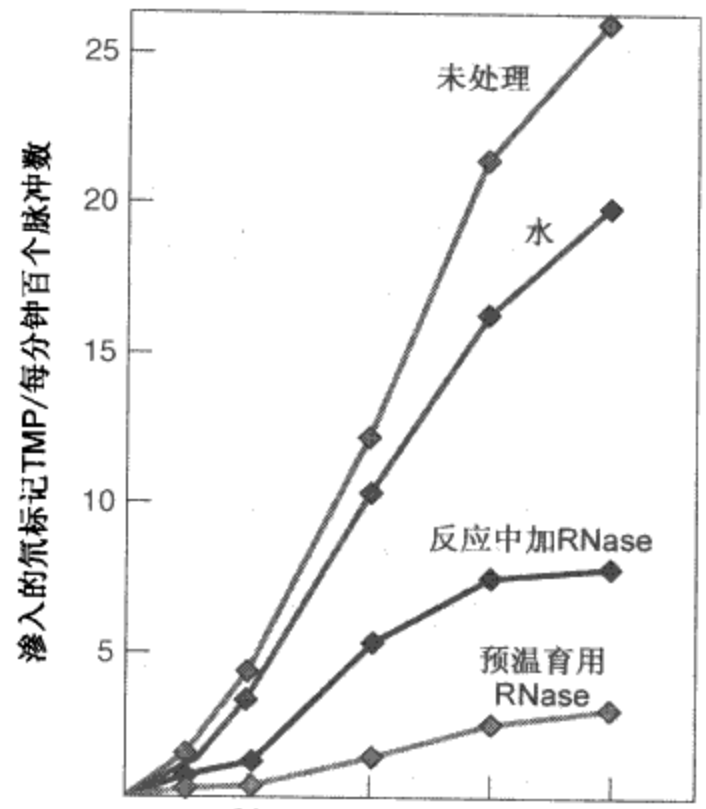


图23.20 RNase对反转录酶活的影响。Baltimore将R-MLV颗粒在不同条件下与4种dNTP（包括[^{3}H] dTTP）一起温育，然后用酸沉淀出产物，用液闪计数器测量产物的放射活性。红色：无处理；紫色：与水进行20min预温育；蓝色：反应体系中加入RNase；绿色：与RNase预温育。（*Source*: Adapted from Baltimore, D., Viral RNA-dependent DNA polymerase. *Nature* 226: 1210, 1970.）

Baltimore检测反应产物的结果表明产物对RNase和碱基水解酶不敏感但对DNase敏感。另外，病毒体只能渗入dNTP，而不能渗入包括ATP在内的核糖核酸。因此，产物的行为与DNA的相似，酶的行为类似于依赖RNA的DNA聚合酶，即反转录酶。Baltimore和Temin用Rous肉瘤病毒颗粒做了类似实验，得出了相似的结果。看起来所有的RNA肿瘤病毒都可能包含反转录酶，并按如图23.18所示的原病毒假说的形式反应。这些已被证明是正确的。

tRNA引物的证据 在分子生物学家开始研究反转录酶的分子生物学机制时，发现病毒的反转录酶和其他已知的DNA聚合酶一样需要引物。1971年，Baltimore及同事用下述方法发现RNA引物接在初始形成的转录物的5′端：用将病毒颗粒与标记dNTP一起温育（同Baltimore和Temin标记法）的方法标记禽成髓细胞瘤病毒（AMV）的初始反转录物。然后用Cs_2SO_4梯度超速离心，通过碱基密度的差异将RNA和DNA分离（RNA比DNA密度大）。

在最初的实验中，Baltimore研究组从病毒颗粒上分离出核酸后迅速进行超速离心。结果如图23.21a，标记DNA的峰值出现在RNA聚集区。这个结果与初始形成的DNA与比其大很多的RNA模板之间仍然通过碱基互补配对结合的假设一致，所以复合体表现得像RNA。如果这个假设成立，那么加热RNA-DNA杂合链将使其变性而释放出DNA产物成为自由分子。当Baltimore和同事做这项实验时，观察到如图23.21b所示的现象：初始DNA产物的密度与纯DNA的密度接近但大一些，像仍然与一些RNA结合在一起。

这种现象可以解释为初始形成的DNA仍然与RNA引物共价结合。为了验证这种可能性，Baltimore及同事用RNase处理初始形成的DNA然后进行超速离心。这次实验结果显示出产物的密度与纯DNA分子的密度完全一致（图23.21c）。因此，新生反转录产物是由RNA为引物所引发的，具体是哪种RNA呢?

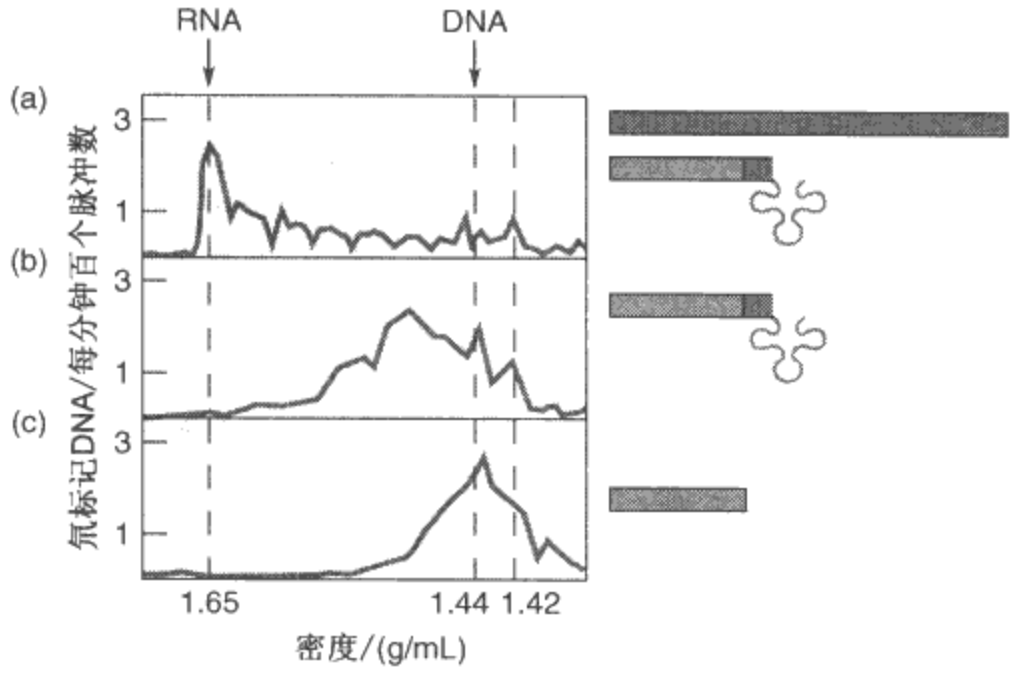


图 23.21　反转录物中含有 RNA 引物。Baltimore 及同事用［^{3}H］dTTP 标记 AMV 颗粒中的反转录物，然后再做如下处理后进行 Cs_2SO_4 梯度超速离心：（a）未处理；（b）加热变性双链多聚核苷酸；（c）通过加热和 RNase 处理消除所有与反转录物结合的引物。右侧图示解释了结果产生的原因。（a）未处理的材料有与 RNA 相近的高密度，因为反转录物很短而与其碱基互补配对结合的病毒 RNA 模板却长得多。（b）加热后的反应体系的密度与纯 DNA 很接近，这是因为 RNA 模板解离，但是仍比纯 DNA 密度大则是因为还与 RNA 引物共价结合。（c）经过加热和 RNase 处理后密度与纯 DNA 一致，因为 RNase 降解了 RNA 引物。上部给出了纯 RNA 和 DNA 的近似密度。（*Source*：Adapted from Verma, I. M., N. L. Menth, E. Bromfeld, K. F. Manly and D. Baltimore, Covalently linked RNA-DNA molecules as initial product of RNA tumor virus DNA polymerase. *Nature New Biology* 233：133，1971.）

在研究反转录病毒颗粒中所有相关分子的过程中，分子生物学家发现了一些 tRNA，其中一种是寄主 $tRNA^{Trp}$ 与病毒 RNA 部分碱基互补配对。这会是引物吗？如果是，它就会与反转录酶结合。为了验证事实是否如此，Baltimore、James Dahlberg 及同事用 ^{32}P 标记寄主 $tRNA^{Trp}$ 和病毒颗粒中的 $tRNA^{Trp}$，将标记的 tRNA 与 AMV 反转录酶混合。混合物用 Sephadex G-100 凝胶过滤（第 5 章）。$tRNA^{Trp}$ 自身进入凝胶，在 #25 孔径有洗脱峰。然而，当寄主和病毒体 tRNA 与反转录酶混合后，与酶一起的洗脱峰在 #20 孔径。因此，反转录酶结合在 $tRNA^{Trp}$ 上。综合以上数据表明 $tRNA^{Trp}$ 是反转录酶的引物。病毒并不编码 tRNA，所以引物必须来自于寄主细胞。

反转录病毒的复制机制　体外反转录的起始产物是一段短的 DNA 片段，称为强终止 DNA（strong-stop DNA）。当我们考虑到病毒 RNA 上与 tRNA 引物结合的位点（引物结合位点，primer-binding site，PBS）时，称之为强终止的原因就显而易见了。距离病毒 RNA 5′端仅为 150 核苷酸（依据反转录病毒），这就意味着反转录酶在到达 RNA 模板末端并终止之前将合成大约 150 核苷酸的 DNA。这就引起了一个让我们感兴趣的问题：接下来会发生什么？

这个问题与反转录病毒复制中的另一个悖论相关，如图 23.22。原病毒比病毒 RNA 长，但是病毒 RNA 又是合成原病毒的模板。特别是病毒 RNA 的 LTR 是不完整的。左侧 LTR 包含一个冗余区（R）和一个 5′非翻译区（U5），右侧 LTR 包含一个 R 区和一个 3′非翻译区（U3）。在其模板缺失左侧 U3 区和右侧 U5 区时，原病毒是如何在其两个末端具有完整的 LTR？Harold Varmus 基于反转录酶还具有 RNase 活性的特性，提出了一种解释：反转录酶固有的 RNase 是 RNase H，它能特异地降解 RNA-DNA 杂合链中的 RNA。

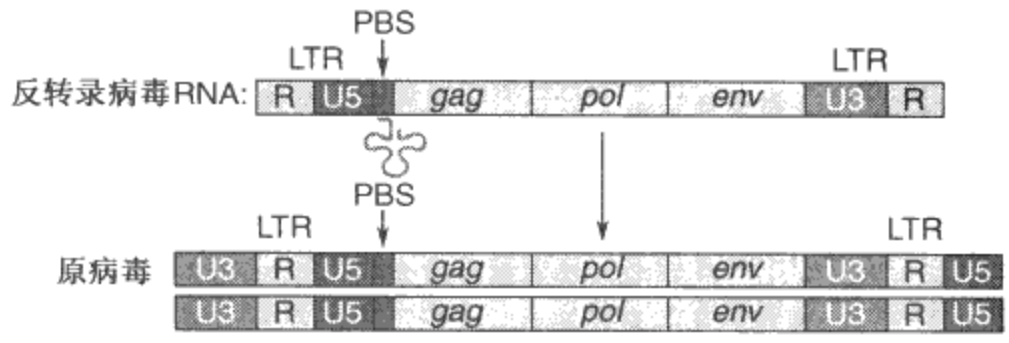


图 23.22　反转录病毒 RNA 和原病毒 DNA 的结构。这是一个非缺陷型反转录病毒 RNA（vrial RNA），包含所有复制所需的基因：外壳蛋白基因（*gag*）、转录酶基因（*pol*）和一个包被蛋白基因（*env*）。另外，它还包含两端的长末端重复序列（LTR），但是重复序列不相等。左侧 LTR 包括一个 R 区、一个 U5 区和这里显示与 tRNA 引物结合的引物结合位点（PBS）。右侧 LTR 包括一个 U3 区和一个 R 区。此外，以病毒 RNA 为模板合成的原病毒 DNA 的两端均含有完整的 LTR（U3、R 和 U5）。

Varmus 的假说如图 23.23 所示。首先，反转录酶利用 tRNA 为引物合成强终止 DNA（图 23.23a）。这一步似乎要到链的终点，但是接下来 RNase H 识别强终止 DNA 和 RNA 模板的杂合链中的 RNA 链，并降解 RNA 中的 R 区和 U5 区（图 23.23b）。降解的 RNA 保留一条 DNA 尾巴（蓝色）通过 R 区与该 RNA 模板的另一末端或其他 RNA 杂交（图 23.23c）。这种与其他 R 区的杂交称为“第一次跳跃”（first jump）。我们假设 DNA 跳到同一 RNA 的另一端，可通过使 RNA 成环完成，而不需要强终止 DNA 从 RNA 左侧末端跳转到右侧末端进行配对。但是 DNA 也可以跳到另一个 RNA 上，而且可能性更大，因为每个病毒颗粒都有两个拷贝的 RNA 基因组。

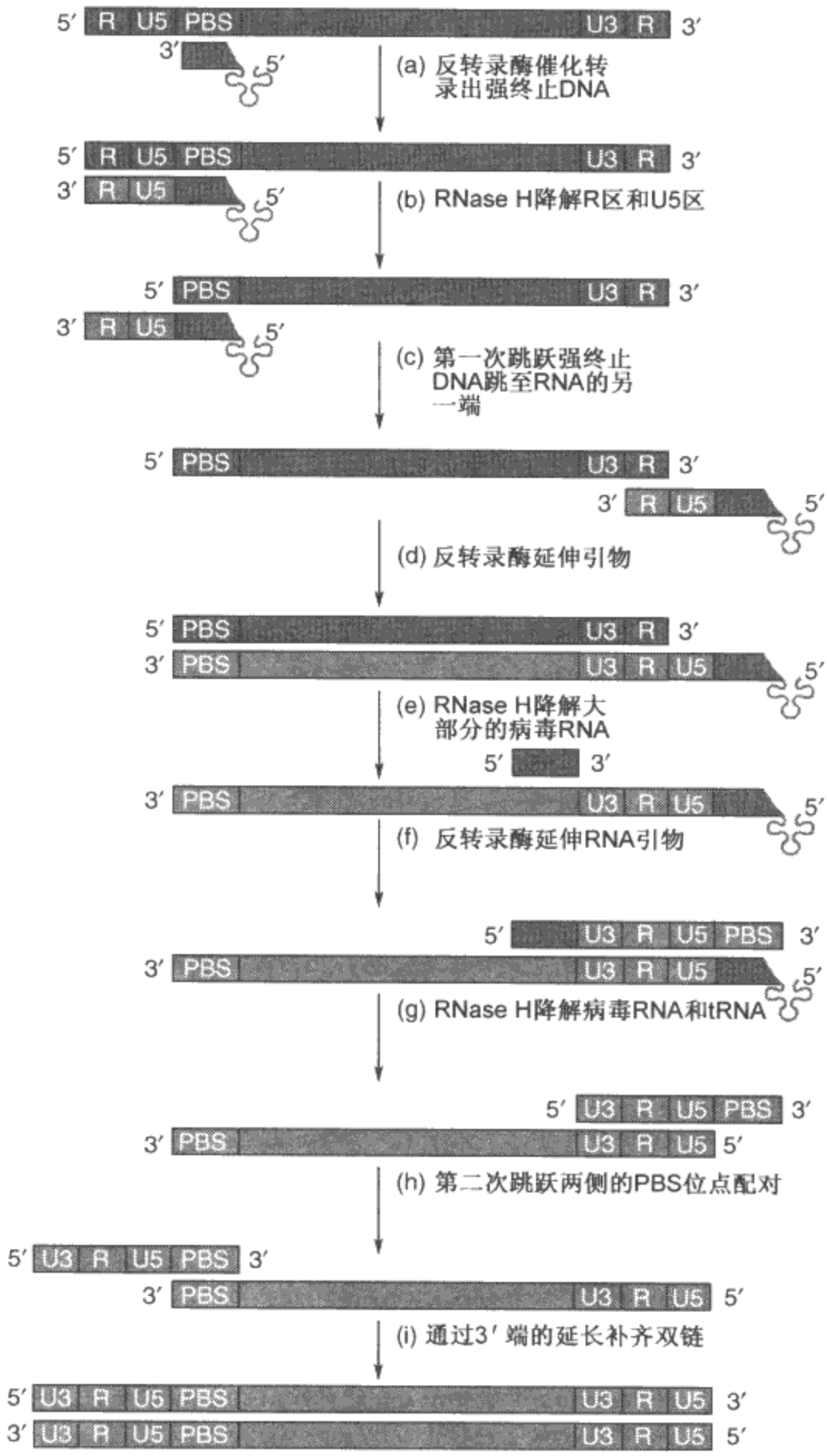


图 23.23　以反转录病毒 RNA 为模板合成原病毒 DNA 的模式图。所有 RNA 用红色表示，DNA 用蓝色表示。带有 3′端能与病毒 RNA 引物结合位点（PBS）杂交的 tRNA 引物用三叶草形状表示。详细步骤介绍见正文。

第一次跳跃以后，强终止 DNA 在模板右末端可以作为反转录酶的引物引导复制余下的病毒 RNA（图 23.23d）。注意，第一次跳跃使得右侧 LTR 补全。从病毒 RNA 的左侧 LTR 复制了 U5 和 R 区，从右侧 LTR 复制了 U3 区。图 23.23e，RNase H降解大部分病毒 RNA，但保留 RNA 的一段小片段与右侧 LTR 连接，作为合成第二链的引物（图 23.23f）。当反转录酶延伸引物到末端，包括 PBS 区，RNase H 降解剩余的 RNA（图 23.23g）作为第二链的引物，它和 tRNA 都与 DNA 配对。这就产生第二次跳跃（图 23.23h），在第二次跳跃中右侧的 PBS 区域与左侧的 PBS 区配对。与第一次跳跃一样，第二次跳跃可以想象成跳到另一个分子上，或者跳到同一分子的另一端。如果跳到同一分子上，周围 DNA 可以环化使得两个 PBS 区碱基配对。第二次跳跃后，后面的工作就是反转录酶的了，利用 DNA 作模板或以末端长的单链突出端作为引物，另一个 DNA 聚合酶合成两条链（图 23.23i）。

一旦原病毒合成，就可以在整合酶（integrase）的催化下插入到宿主基因组。这种酶原本是 *pol* 基因编码的多聚蛋白（polyprotein）的一部分，*pol* 基因还编码反转录酶和 RNase H。整合酶是多聚蛋白经蛋白酶水解而成，蛋白酶也是同一多聚蛋白的一部分。蛋白酶可以从多聚蛋白上水解出自身（值得注意的是，多数有希望抵抗 AIDS 的药物都是 HIV 蛋白酶的蛋白酶抑制基因）。一旦原病毒整合到宿主基因组，它将由寄主 RNA 聚合酶 II 转录产生病毒 RNA。

小结　反转录病毒通过一个 RNA 中间体进行复制。当反转录病毒感染细胞后，利用病毒编码的反转录酶产生一个自身的 DNA 拷贝完成 RNA→DNA 的反应，RNase H 降解复制过程中产生的 RNA-DNA 杂合链中的 RNA。寄主 tRNA 作为反转录酶的引物。最终合成的病毒 RNA 的双链 DNA 拷贝随后插入寄主基因组，可以被寄主聚合酶 II 转录。

反转录转座子

所有的真核生物都能从一个 RNA 中间体复制转座子，因此有赖于反转录酶。这些反转录转座子根据复制模式不同分为两类。第一类是带有 LTR 的反转录转座子，复制模式与反转录病毒很相似，但是它们不会在细胞间传递病毒颗粒。毫无疑问，这些转座子被称为含有 LTR 的反转录转座子（LTR-containing retrotransposon）。第二类是非 LTR 的反转录转座子（non-LTR retransposon）。

含有 LTR 的反转录转座子　首次在果蝇（*Drosophila melanogaster*）和酵母（*Saccharomyces cerevisiae*）中发现了反转录转座子。果蝇转座子的原型称为 *copia*，因为它在基因组中有极大的数量（copious quantity）而得名。事实上，*copia* 及其称为 *copia* 类似因子的相关转座子占果蝇全基因组的 1%。酵母中与 *copia* 类似的转座因子称为 Ty，是酵母转座子（transposon yeast）的简称。这些转座子中的 LTR 与反转录病毒中的 LTR 相似，表明它们的转座类似于反转录病毒的复制。实际上，许多证据表明事实果真如此。以下总结出从 RNA 中间体复制 Ty1 因子与反转录病毒的复制相似的证据。

1. Ty1 编码一个反转录酶。Ty 中的 *tyb* 基因编码一个蛋白质的氨基酸序列与反转录病毒中 *pol* 基因编码的反转录酶相似。如果 Ty1 因子确实编码一个反转录酶，那么在 Ty1 诱导转座时便会出现这个酶；而且，*tyb* 突变将阻断反转录酶的出现。Gerald Fink 及同事设计了相关实验并得到了预想的结果。

2. 全长 Ty1 RNA 和反转录酶活性均和与反转录病毒颗粒相似的颗粒连接在一起。这些颗粒只在诱导 Ty1 转座的酵母细胞中存在。

3. Fink 及同事在一个巧妙的实验中将一个内含子插入到 Ty1 因子中，在转座之后再次分析这个因子。内含子不见了！这个发现与细菌转座类型相悖，细菌中转座的 DNA 与其亲本相似。但是随后的机制相同（图 23.24）：首先转录 Ty 因子，随后是内含子和其他因子 RNA 剪切掉内含子；最后，剪切后的 RNA 在一个类病毒颗粒中反转录，在酵母基因组的一个新位点重新插入。

4. Jef Boeke 及其研究小组证实寄主 $tRNA_i^{Met}$ 作为 Ty1 反转录的引物。首先，他们在与寄主 $tRNA_i^{Met}$ 互补的 Ty1 因子的 PBS 的 10 个核苷酸中突变 5 个。突变体阻断转座的发生，可能是由于这些改变让 tRNA 引物无法结合到 PBS 上。然后 Boeke 及同事对一个拷贝的寄主 $tRNA_i^{Met}$ 基因的突变进行了修复，并使 $tRNA_i^{Met}$ 能重新与突变的 PBS 结合。这些突变体使突变的 Ty1 因子的转座活性得以恢复。正如在本书中经常提到的那样，这种突变抑制是两个分子相互作用重要性的强有力证据。在此例中，两个分子就是 tRNA 引物和 Ty1 因子上的结合位点。

Copia 及其类似因子分别具有这些 Ty 的特征，也和 Ty 的转座方式相同。人类也含有 LTR 的反转录转座子，但是缺少功能基因 *env*。最显著的例子是人类内源性反转录病毒（human endogenous retroviruse，HERV），它占人类基因组的 1%～2%。迄今为止，还没发现能完成转座的 HERV，所以 HERV 可能是原病毒反转录转座的残留物。

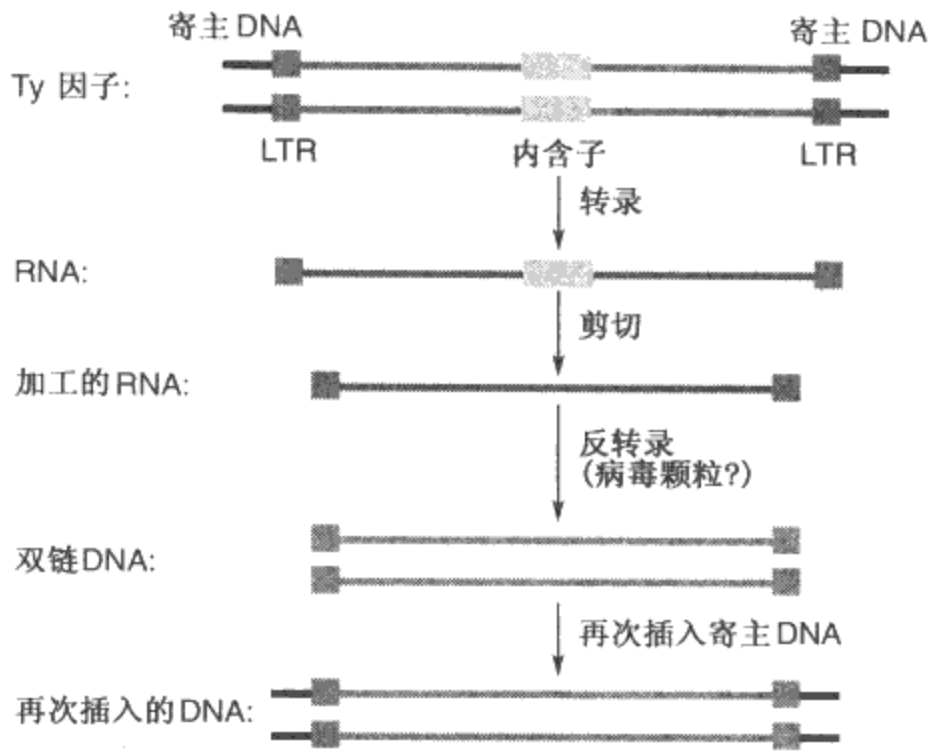


图 23.24　Ty 的转座模式。用实验手段在 Ty 因子中插入一个内含子（黄色）。Ty 因子转录产生一个有内含子的 RNA 拷贝。这个转录物发生剪切后，RNA 可能在一个类病毒颗粒中被反转录。形成的双链 DNA 再次插入酵母基因组。

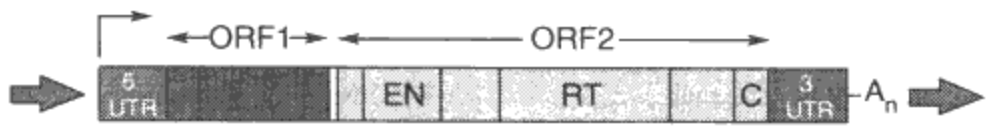


图 23.25　L1 因子示意图。ORF2（黄色）的亚区有 EN（内切核酸酶）、RT（反转录酶）和 C（半胱氨酸富集区）。两端的紫色箭头表示寄主 DNA 的正向重复序列，右端的 A_n 表示多聚 A [poly（A）]。

小结　一些真核生物转座子，包括酵母中的 Ty 和果蝇中的 *copia* 转座的方式是相似的。它们以寄主基因组 DNA 起始，产生一个 RNA 拷贝，然后可能在一个类病毒颗粒中反转录，最后插入基因组的新位点。HERV 可能以相同的方式转座，但绝大多数丧失了转座能力。

非 LTR 的反转录转座子　至少在哺乳动物中，缺失 LTR 的反转录转座子比含有 LTR 的反转录转座子要多得多。其中丰度最高的是长散布因子（long interspersed element，LINE），其中 L1 至少存在100 000拷贝，占人类基因组的 17%，尽管 97% 的 L1 拷贝缺失 5′端且绝大多数（除了 60～100 拷贝）都有阻碍转座的突变。L1 因子的普遍存在意味着这种被传统分类归于“垃圾 DNA”的反转录转座子在基因组中的比重是人类所有外显子的 5 倍。图 23.25 是 L1 因子的完整示意图，注明了它的两个 ORF。ORF1 编码一个 RNA 结合蛋白（p40），ORF2 编码一个具有两种活性的蛋白质，分别是内切核酸酶活性和反转录酶活性。和这类反转录转座子一样，L1 也是聚腺苷酸化的。

正如我们所见，LTR 对含有 LTR 的反转录转座子的复制是至关重要的，那么非 LTR 反转录转座子是如何复制的呢？尤其是它们的引物是什么？答案是：它们的内切核酸酶在靶 DNA 上产生一个单链断裂，反转录酶便利用这个新形成的 DNA 3′端作为引物进行反转录。Thomas Eickbush 及同事对 R2Bm 的研究为这一机制提供了最好的证据，R2Bm 是线虫（*Bombyx mori*）中一种 LINE 类似因子。这个因子和哺乳动物的 LINE 一样编码反转录酶，但是与 LINE 不同的是它有特定的靶位点——寄主的 28S rRNA 基因。这种机制使我们易于研究插入机制。

Eickbush 研究小组首先揭示了单一 ORF 的 *R2Bm* 编码内切核酸酶，它能特异地切割 28S rRNA 靶位点。随后，他们纯化了这个内切核酸酶（和一个活性必需的 RNA 辅助因子），将它加入含有靶位点的超螺旋质粒中。如果质粒的一条链被切开，超螺旋质粒将转变成松弛的环状。如果两条链都被切开，将形成一个线性 DNA。图 23.26a 和图 23.26b 的结果显示，实验中很快就出现了开环，接着开环慢慢转变为线状 DNA。因此，R2Bm 内切核酸酶能快速地在靶位点切开 DNA 的一条链，然后用较慢的速度切开另一条链。切割是特异的，当质粒缺少靶位点时，核酸酶不能切割其中的任意一条链。

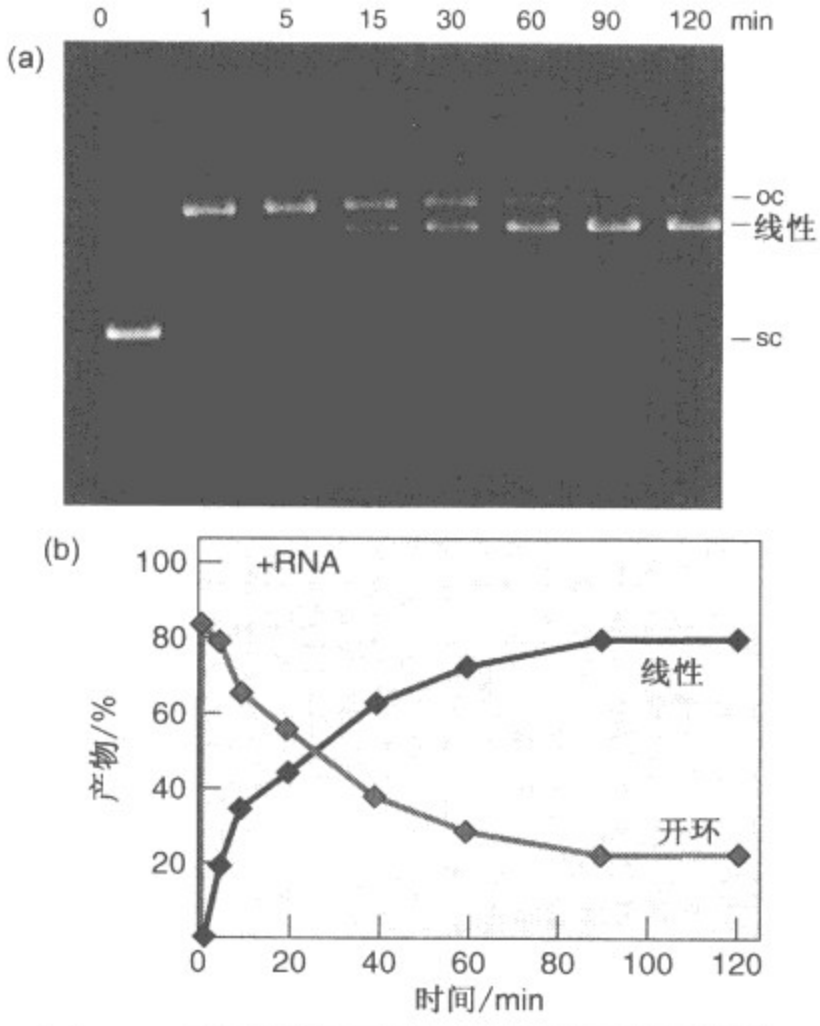


图 23.26　DNA 切口和 R2Bm 的内切核酸酶活性。Eickbush 研究小组将有 R2Bm 反转录转座子靶位点的超螺旋质粒和纯化的 R2Bm 内切核酸酶混合，加入或不加入 RNA 辅助因子，然后对质粒进行电泳看其是否产生切口（松弛形成开环形式）或切开两条链产生线状 DNA。（a）电泳凝胶用溴化乙锭染色。右侧标明了超螺旋（sc）质粒、开环（oc）质粒和线性（linear）质粒的位置；（b）图表示意图（a）的结果；（c）相同的实验但是不加入 RNA 共因子的实验结果。（*Source*：Adapted from Luan, D. D., M. H. Korman, J. L. jakubczak and T. H. Eickbush, Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target sige: a mechanism for non-LTR retrotransposition. *Cell* 72 (Feb 1993) f. 2, p. 597.)

接着，研究者们除去 RNA 辅助因子，结果发现蛋白质仍可依靠自身快速切割靶位点的单链，但是检测不到另一条链被切开（图 23.26c）。他们还观察到线型 DNA 通过 T4 DNA 连接酶发生自身环化，这个反应需要 5′-磷酸基团。因此，R2Bm 内切核酸酶剪切会产生 5′-磷酸和 3′-羟基。然后，利用内切核酸酶形成单链切口并通过引物延伸分析发现转录链有切口（转录链的切口阻断了 DNA 聚合酶在引物上的延伸，但是另一条链的引物延伸没有受到切口的阻碍）。通过对两条链 DNA 切口上更精确的引物延伸实验，他们准确的定位了切口位点并发现两条链被切掉 2 个碱基对。

为了确定被切割的靶 DNA 是否真的作为引物分子，Eickbush 及同事设计了一个体外反应体系，以一段短的预先切开的靶 DNA 为引物，R2Bm RNA 为模板，加入 R2Bm 反转录酶和 4 种 dNTP（包括 [^{32}P] dATP）。产物电泳后放射自显影，检测产物是否为预期的大小。图 23.27a 说明了在分子水平上发生的变化，结果如图 23.27b。当模板是非特异性 RNA 时，反应没有产物（泳道 1），若加入 R2Bm RNA 将产生一条 1.9kb 的亮带。这是预期的结果吗？我们很难得知，因为我们不知道反转录酶确切地移动了多远，而且得到的还是有些小支链的多聚核苷酸，但是可以推测已经很接近了，因为引物长 1kb，模板长为 802nt。为了进一步研究产物的特性，Eickbush 及同事在反应体系中加入双脱氧 CTP（泳道 3）。正如预期的一样，它引起了反转录在多个位点的提前终止，形成一条模糊带。在另一个反应体系中，于电泳之前用 RNase A 处理产物降解模板上没有与反转录物碱基互补配对的部分。泳道 4 显示的就是长度缩短为 1.8kb 的产物带，表明 RNA 模板 5′端约 100nt 降解，所以反转录酶在多数

情况下都没有完成它的任务。研究者们还在电泳前用 RNase H 处理产物，得到了一条约 1.5kb 的弥散带。RNA 模板在这个过程中被降解了，因为它和产物杂交在一起。仍有比 1kb 长的泳带表明 DNA 的一条链延伸了。泳道 6 是另一个阴性对照，用非特异 DNA 代替靶 DNA。

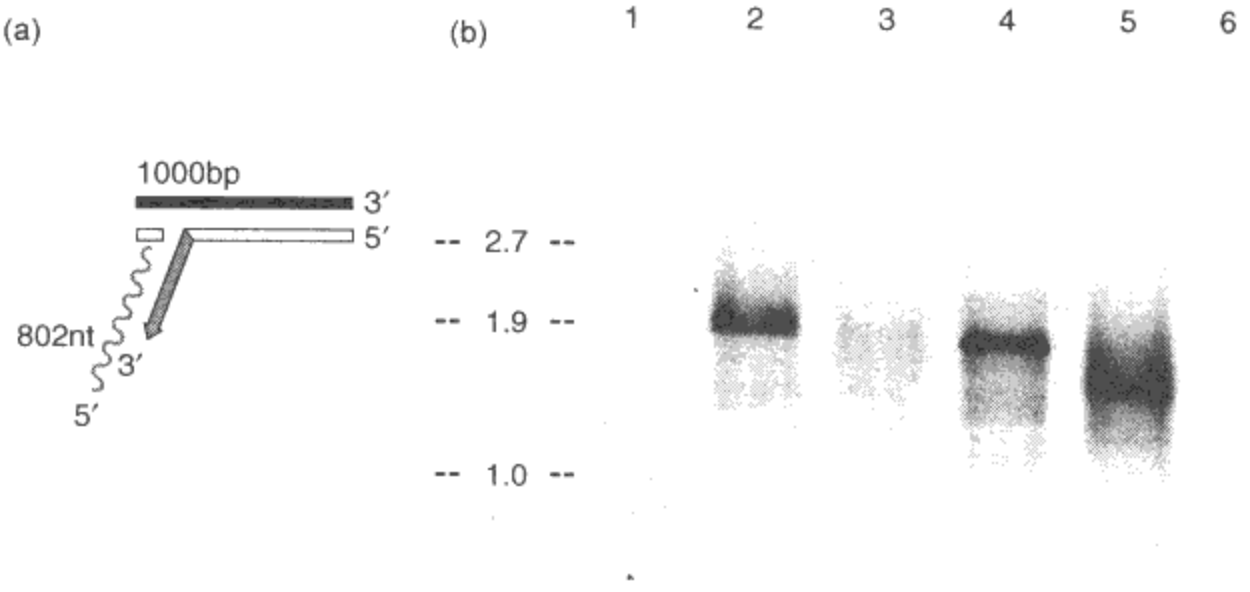


图 23.27　R2Bm 反转录中靶-引物的证据。（a）R2Bm 内切核酸酶在长为 1kb 的靶 DNA 靠近左侧末端形成切口并利用新的 3′端启动 802nt 转座子 RNA 的反转录的预期产物模型。反转录物（蓝色）与引物（黄色）共价结合。底部 DNA 链的左侧也着上黄色。相对的 DNA 链用黑色表示。（b）实验结果。Eickbush 和同事们首先制备了靠近左侧末端具有靶位点的长 1kb 靶 DNA。加入 R2Bm RNA、ORF2 产物和 dNTP，包括用于标记反转录物的［^{32}P］dATP。然后将产物电泳并放射自显影。泳道 1：用非特异 RNA 替代 R2Bm RNA；泳道 2～6：用 R2Bm RNA；泳道 3：在反转录反应体系中加入双脱氧-CTP；泳道 4：电泳前用 RNase A 处理产物；泳道 5：电泳前用 RNase H 处理产物；泳道 6：用非特异靶 DNA。（*Source*：Luan，D. D.，M. H. Korman，J. L. Jakubczak and T. H. Eickbush，Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site：a mechanism for non-LTR retrotransposition *Cell* 72（Feb 1993）f. 4，p. 599. Reprinted by permission of Elsevier Science.）

对更倾向朝左侧延伸的靶 DNA（靶位点在中间）进行了类似的实验，出现了一个明显的大写“Y”字形产物（如图 23.27 预期），表明反转录在第二链剪切之前发生。如果第二链的剪切先发生，产物会更小且呈线型。为了证明靶 DNA 是引物，Eickbush 小组设计了一个 PCR 反应，用靶 DNA 与反转录物杂交作为引物，得到了预期大小和序列的产物。

综合以上和其他数据，H. H. Kazazian 和 John Moran 提出了 L1 转座模型（如图 23.28）。首先转录出转座子，转录物加工。mRNA 离开细胞核到细胞质中翻译。它联合自身的两个产物——p40 和 ORF2 产物重新进入细胞核。在细胞核中，ORF2 产物的内切核酸酶活性将靶 DNA 切开一个切口。对于 L1 来说，DNA 的任意区域都可以是靶位点。然后，ORF2 产物的反转录酶活性利用靶 DNA 上由内切核酸酶产生的 3′端为引物拷贝 L1 RNA。这种机制因此被称为靶引发的反转录转座（target-primed retrotransposition）。最后，进行一些我们尚不清楚的步骤，形成 L1 的第二链，靶位点的第二链发生剪切，L1 因子连接到一个新的宿主中。

从本节开头我们了解到 L1 因子占人类基因组的 17%。随后我们即将看到，这些因子在转座时可以携带因组 DNA 片段。所以，估计 L1 因子直接和间接地构成人类基因组的 30%。而且，在植物和动物中都发现了类 L1 因子。因此，这些因子很古老，至少有 6 亿年。由于在经过 2 亿年的进化后，相同的 DNA 序列互相之间丢失的片段也相同，人类基因组中 L1 因子所占比例实际上大约是 50%。

你会怀疑像 L1 这样在人类基因组中广泛存在的因子一定有副作用，事实上也发现了许多能引起人类疾病的 L1 中间体突变。特别是，在引起血友病的凝血因子 VIII 基因、引起杜兴氏肌营养不良的 *DMD* 基因和辅助产生一种结肠癌——大肠息肉瘤的 *APC* 基因中都发现了 L1 的拷贝。在最后一种情况中，患者癌细胞的 *APC* 基因中含有 L1 因子，但是正常细胞中没有。因此这种转座是患者的一种体细胞突变。

更让人吃惊的是 L1 因子同样也有有利的一面。能是端粒酶的来源，端粒酶可以维持我们染色体的末

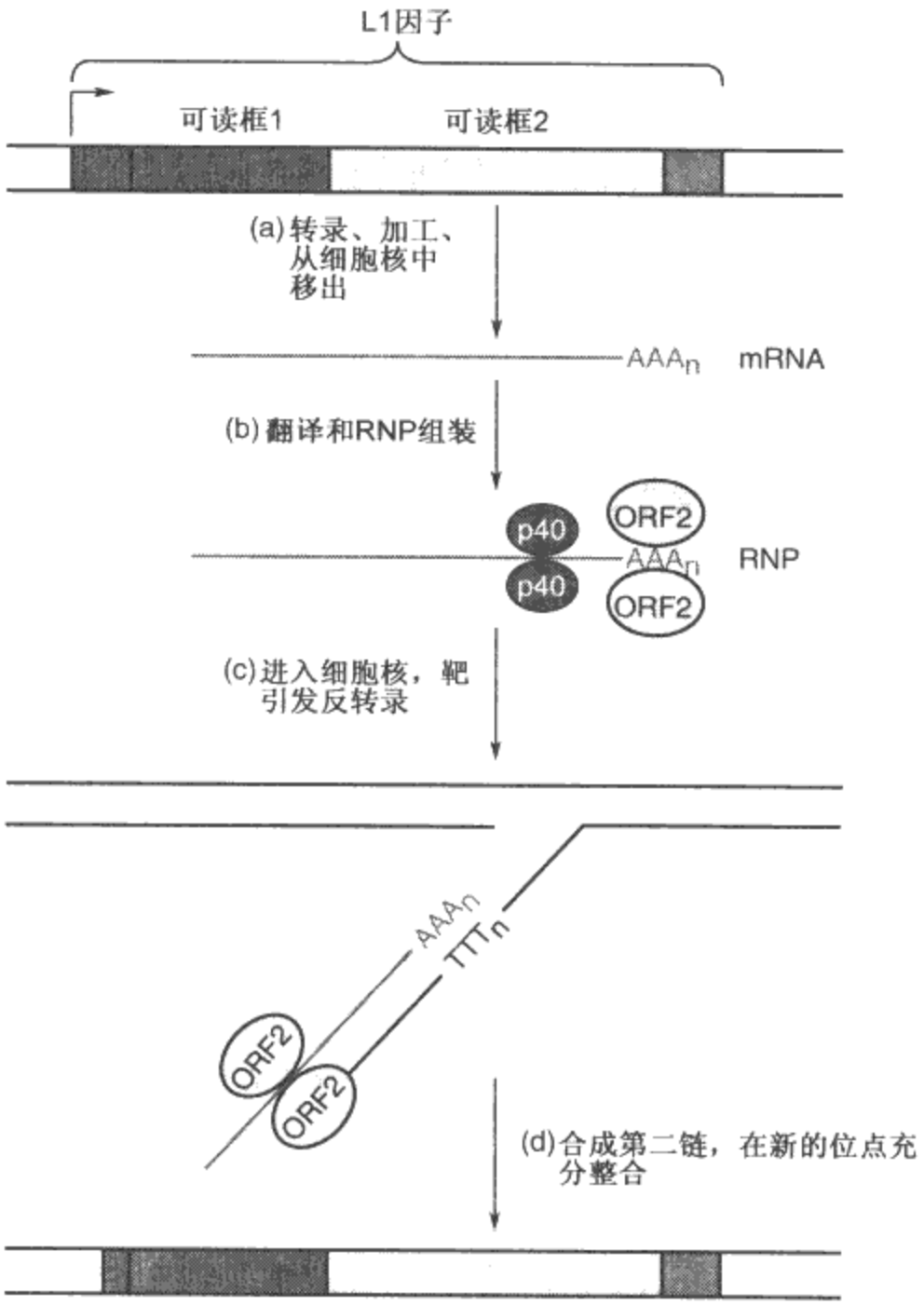


图 23.28　L1 转座模型。（a）L1因子被转录、加工、从细胞核中移出。（b）mRNA 翻译产生 ORF1 产物（p40）和 ORF2 产物，具有内切核酸酶活性和反转录酶活性。这些蛋白质与 mRNA 结合形成一个 RNP。（c）核糖核蛋白重新进入细胞核。内切核酸酶切开靶 DNA（基因组的任何位点），反转录酶利用新的 DNA3′端为引物合成反转录物。（d）经过一系列尚不明了的步骤，L1 第二链形成，组成因子，通常截去 5′端连接到靶 DNA。

端（尽管反面也是真的）。最无法确定其利弊特性的是 L1 利于外显子改组，使得基因间外显子交换。之所以发生这种情况是因为 L1 的聚腺苷酸化信号很微弱，所以聚腺苷酸化机制常常避开寄主转录物下游部分的多聚腺苷酸化位点。这种途径的 RNA 聚腺苷酸化包含一段与 L1 RNA 连接的人类 RNA，这段人类 RNA 将与 L1 因子整合随后进入的反转录物中。这种结合有时是有害的，但有时能在原有的基因中产生新的基因，能产生新的蛋白质和有用性状。

为什么 L1 因子的聚腺苷酸化信号微弱呢？Moran 提出了如下解释：如果聚腺苷酸化信号很强，人类基因中插入这些因子将引起聚腺苷酸化转录物的提前形成，那么下游所有的外显子都将丢失。这可能钝化基因甚至引起寄主死亡。与反转录病毒不同，一方面，L1 因子不能从一个个体转移到另一个个体，而是在寄主中生存直到死亡；另一方面，微弱的聚腺苷酸化信号使得这些因子插入到人类基因的内含子中而不会打断这些基因的多数转录物。因此，由于内含子 DNA 量远远多于外显子的，L1 因子可以在人类基因组的很大区域相对安全地拓展。

小结　LINE 和类 LINE 因子是缺乏 LTR 的反转录转座子。这些因子能编码使靶 DNA 产生切口的内切核酸酶。这些因子还有利于新的 DNA 3′端引发 RNA 反转录。第二链合成后，因子在其靶位点复制。当 LINE 转录时开始新一轮转座。因为 LINE 聚腺苷酰化信号微弱，所以 LINE 的转录经常包括一个或者更多宿主 DNA 的下游外显子。

非自主反转录转座子　另一类非 LTR 反转录转座子（nonautonomous retrotransposon）成员不编码蛋白质，所以它们和 LINE 这样能胜任转座的自主转座子不同。相反，它们依赖其他因子——可能是广泛存在的 LINE 提供给它们需要的包括反转录酶在内的蛋白质来实现转座。研究最深入的非自主反转录转座子是 Alu 因子，之所以称为 Alu 因子是因为含有能被限制酶 *Alu* I 识别的 AGCT 序列。长约 300bp，在人类基因组中有上百万个拷贝。因此，它们会比 LINE 更为成功。成功的一个原因可能是 Alu

因子的转录物含有类似 7SL RNA 的结构域，这种结构域包含有助于核糖体结合到粗糙内质网的信号识别颗粒的正常部分。两个信号识别颗粒蛋白与 Alu 因子 RNA 紧密结合，携带它到核糖体，在核糖体上翻译 L1 RNA。这就将 Alu 因子 RNA 放在合适位置上，以至于使其能和所需要的蛋白质结合，从而被反转录并插入到新的位点。由于该因子很小，所以 Alu 因子及其类似因子都被称为短散布因子（short interspersed element，SINE）。

LINE 还可能在人类基因组结构中起作用，便于产生被加工的假基因（processed pseudogene）。通常假基因是与正常基因序列相似的 DNA，但由于这样或那样的原因没有功能。有时它们内部有翻译终止信号；有时它们的剪切信号钝化或丢失；有时它们的启动子钝化。通常这些问题结合起来就阻止了基因的表达。它们显然来自基因复制及其随后突变的累积。这种过程对寄主是无害的，因为原有的基因保留了功能。

被加工的假基因也来自基因复制，但是显然是通过反转录途径。我们十分怀疑 RNA 是加工型假基因形成的中间体，因为：①这些假基因通常有一段短的 poly（A）尾巴，可能来自 mRNA 的 poly（A）尾巴；②被加工假基因缺乏内含子，但它们的原始基因通常都有。同 Alu 因子并不是来自 mRNA 一样，LINE 为 mRNA 反转录并插入寄主基因组的过程提供了分子机制。

小结 非自主反转录转座子包括人类中丰富的 Alu 因子和脊椎动物中的类似因子。因为它们不编码任何蛋白质所以其自身不能发生转座。相反它们可以利用其他因子（如 LINE）的反转录转座机制。被加工假基因可能来自相同的途径：通过 LINE 机制反转录 mRNA 后插入基因组。

II 型内含子 第 14 章中我们了解了 II 型内含子存在于细菌线粒体和叶绿体基因组，是自我剪切内含子，能形成套索中间体。1998 年，Mariene Belfort 的研究小组发现一个特定基因中的 II 型内含子可以插入到基因组其他地方的无内含子型的相同基因上，这个过程称为**反转录归巢**（retrohoming），其机制如图 23.29 所示。基因携带的内含子首先被转录，然后内含子突出形成套索后被剪切。这个内含子可以被同一基因的无内含子形式识别，而后通过逆剪接的方式整合进去。反转录形成内含子的一个 cDNA 拷贝，合成的第二链的 DNA 取代 RNA 内含子。

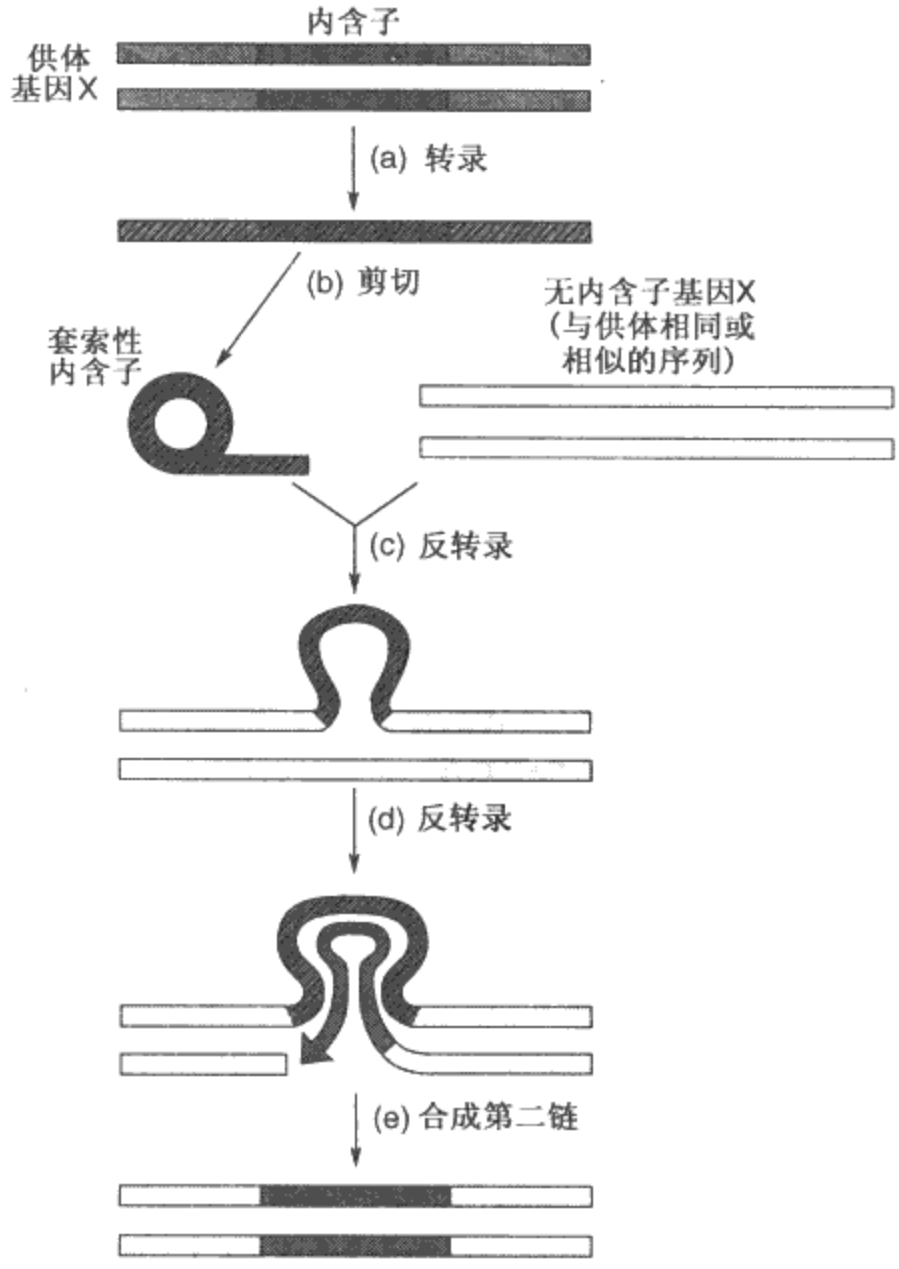


图 23.29 反转录归巢。（a）携带有 II 型内含子（红色）的供体基因 X（蓝色）转录产生 RNA（RNA 用阴影表示）。（b）转录物剪接产生套索型内含子（lariat-shaped intron）。（c）内含子自身逆剪接为 X 基因的另一拷贝，两者有相同或相似的序列只是剪接后代片段不含内含子。（d）编码反转录酶的内含子以底部 DNA 链的切口为引物合成内含子的一个 DNA 拷贝。箭头所指反转录物的 3′端。（e）顶部链的 RNA 内含子移位合成内含子的第二链（DNA 形式）。完成了反转座至原位的过程。

1991 年，Phillip Sharp 提出 II 型内含子可能是现代剪接体内含子的前身，部分原因是两者的剪接机制很相似。2002 年，Belfort 和同事们揭示了剪接是如何发生的。他们发现了一个细菌 II 型内含子的真正的反转录转座而不仅是所谓的反转录归巢。因此，这个内含子可转移到了一系列新的位点，而不仅是插入到这个内含子同源基因的无内含子拷贝上。

为了发现反转录转座，Belfort 和同事们构建了一个含有 *Lactococcus lactis* L1. LtrB 修饰的 II 型内含子的质粒，在反方向含有一个卡那霉素抗性基因，并被可自我剪接的 I 型内含子中断。为了能表达卡那霉素抗性，II 型内含子最先被转录，因此清除中断的 I 型内含子。然后，转录物被反转录成 DNA，该 DNA 能插入宿主 DNA，随后被正向转录而非反方向。只要 II 型内含子保留 RNA 形式时，它并不编码卡那霉素抗性基因，因为它的抗性基因已在反方向上被转录成反义 RNA 了。

在 Belfort 小组选择卡那霉素抗性细胞的时候，他们发现转座发生相对较少，但是确实有一定的频率。这种转座的一个有趣的特点是多数都发生在 DNA 复制的后随链上。这一发现说明转座现象在复制过程中发生，并以一段后随链生成的短的 DNA 片段（第 20 章）作为引物引导完成诸如 L1 转座程序（图 23.28）中的一类靶引导反转录。注意，这种机制要求转座子和靶 DNA 之间没有同源性，这样复制后随链才能在基因组的任意位点产生切口。

一旦 II 型内含子发生了反转录转座子，将保留自我剪接的能力，因此靶基因也能保留功能。这样，II 型内含子在现代真核细胞的前体中能更容易更安全地进行增殖。最后，真核细胞中出现了一些能使剪接更有效的剪接体。

小结 II 型内含子通过 RNA 内含子插入到基因内反转录归巢到同基因的无内含子拷贝中，接着反转录和合成第二链。II 型内含子也可以 DNA 后随链片段为引物，通过一个不相关基因中插入的 RNA 内含子发生反转录转座。II 型内含子的这种反转录转座可能是现代真核细胞剪接体内含子的祖先，或许能解释其在高等真核细胞中的广泛出现的现象。

总结

转座因子或转座子是可以在不同位置间移动的一段 DNA。一些转座因子复制后在原先位置保留一个拷贝后在新位置加入一个拷贝；另一些转座子不复制直接从原位置移动。细菌的转座子类型包括：①插入序列，如 IS1，仅保留了转座必需的一个基因，两侧翼有反向末端重复序列；②转座子，如 Tn3，与插入序列类似但是至少含有一个额外基因，这个基因通常具有抗生素抗性。

真核生物的转座子采用更多的复制策略，如玉米中的 *Ds/Ac* 和果蝇中的 P 因子等 DNA 转座子的转座与细菌中的 DNA 转座子（如 Tn3）的转座相似。

哺乳动物的免疫球蛋白基因甚至利用类似的转座机制进行重排。脊椎动物免疫系统在它们生成的免疫球蛋白中产生巨大的多样性。多样性的主要来源是异质池选择的 2～3 个基因部分的组装。这种基因片段的组装称为 V（D）J 重组。V（D）J 重组的重组信号序列（RSS）由被一段 12bp 或 23bp 片段隔开的一个七聚体和一个九聚体组成。重组仅发生在 12 信号和 23 信号之间以确保重组基因中仅有一个编码区。在人类 V（D）J 重组中 RAG1 和 RAG2 起关键作用。它们在毗连 12 信号或 23 信号的 DNA 中引入一个单链切口。在新形成的 3′-羟基处发生酯基转移反应攻击相反的链，使其在编码片段末端断裂形成发夹结构。随后发夹结构断裂，以非精确方式相互链接，通过增、减碱基连接编码区。

反转录转座子有两种不同的类型。含有 LTR 的反转录转座子的复制与反转录病毒的相似性表现在下列方面。反转录病毒通过 RNA 中间体复制。当一个反转录病毒感染细胞，利用病毒编码的反转录酶完成 RNA →DNA 的反应，产生一个自身 DNA 拷贝，RNase H 降解在复制过程中产生的 RNA-DNA 杂合链中的 RNA。寄主 tRNA 作为反转录酶的引物，形成病毒 RNA 的双链 DNA 拷贝插入寄主基因组，利用寄主聚合酶 II 转录。酵母中反转录转座子 Ty 和果蝇中 *copia* 的复制过程相同。寄主基因组 DNA 首先形成 RNA 拷贝，然后可能在一个类病毒颗粒中反转录产生 DNA，可以插入到新的位置。

非 LTR 反转录转座子为另一类真核细胞反转录转座子，以不同的方式引发反转录。例如，LINE 及 LINE 类似因子编码切割靶 DNA 的内切核酸酶。然后该因子利用新的 DNA 3′端引发因子 RNA 的反转录。第二链合成后，这个因子在其靶位点复制。当 LINE 转录时新一轮的转座开始。因为 LINE 的聚腺苷酸化信号微弱，LINE 的转录通常包括了寄主 DNA 下游的一个或两个外显子；这有助于将寄主外显子转运到基因组的新位置。

非自主性的非 LTR 反转录转座子包括广泛存在于人类的 Alu 因子和其他脊椎动物的类似因子。由于它们不编码任何蛋白质，自身不能发生转座。但它们利用如 LINE 等其他因子的反转录转座机制。被加工型的基因可能源于相同的途径：mRNA 可能

以 LINE 机制反转录后插入到基因组。

II 型内含子代表存在于细菌和真核细胞中的另一类非 LTR 反转录转座子。它们可以在反转录和第二链合成后，在一个基因中插入一段 RNA 内含子反转录归巢到同一基因的无内含子拷贝中。II 型内含子还能以 DNA 后随链的片段为引物，在靶位点引发反转录的方式是，在一个非相关基因中插入一个 RNA 内含子引发反转录转座。II 型内含子的反转录转座可能为现代真核细胞剪接体内含子提供了祖先，并解释其在高等真核细胞中大量出现的现象。

复习题

1. 描述证明细菌转座子中包含末端反向重复序列的实验及其结果。

2. 比较细菌的转座子 IS1、Tn3 和真核生物 *Ac* 转座子的遗传图。

3. 先简单后具体地描绘 Tn3 的转座机制。

4. 描绘保留式转座机制。

5. 解释转座导致玉米产生斑点籽粒的原因。

6. 描绘抗体蛋白的草图，显示轻链和重链。

7. 解释成千免疫球蛋白基因产生上百万种抗体蛋白的原因。

8. 图解在 β 淋巴细胞成熟过程中产生的免疫球蛋白轻链和重链基因间的重排。

9. 说明 V（D）J 连接信号怎样确保在一个成熟重排基因中仅含免疫球蛋白基因中的一个。

10. 绘制检测七聚体、九聚体和重组信号序列上间隔区重要性的报告质粒图，并说明如何用该质粒检测重组。

11. 提出免疫球蛋白基因重组信号序列处 DNA 链的剪切和重接模式。这种机制对抗体多样性有何贡献？

12. 叙述体外实验并给出结果，证明免疫球蛋白重组信号序列的剪切导致发夹结构形成。

13. 提出反转录病毒颗粒中反转录酶活性及 RNase 影响该酶活性的证据。

14. 叙述实验并给出结果，证明反转录病毒中强终止反转录物与 RNA 基因组碱基互补配对并与 RNA 引物共价附着。

15. 阐明反转录病毒基因组 RNA 中 LTR 与反转录病毒原病毒中 LTR 的结构差异。

16. 图示反转录病毒 RNA 到原病毒的转变过程。说明这将如何解释问题 15 中提及的差异。

17. 比较反转录病毒的复制与反转录转座子的转座。

18. 小结反转录转座子通过 RNA 中间体转座的证据。

19. 叙述实验并给出结果，证明 LINE 类似因子的内切核酸酶特异切割该因子靶 DNA 的一条链。

20. 叙述实验并给出结果，证明 LINE 类似因子利用靶 DNA 的切割链作为该因子反转录引物。

21. 提出一个 LINE 类似因子反转录转座的模式。

分析题

1. 某种转座子的转座酶在宿主 DNA 上 5bp 间距处错位切割。这会对在转座子插入周围的宿主 DNA 产生什么结果？图解错位切割对宿主 DNA 的影响。

2. 设计一个实验检测一个假设的携带有两个抗生素抗性基因 *Stealth* 转座子从一个质粒转移到另一个携带有氯霉素抗性基因的质粒的转移速率（*Stealtb* 携带氨苄青霉素抗性）。

3. 鉴定 Tn3 转座子在下列基因中发生突变后进行不彻底转座的末端产物：

a. 转座酶；

b. 解旋酶。

4. A 质粒中的 TnT 转座子转座到 B 质粒。在共联体中 TnT 的拷贝数有多少？相对两个质粒而言，TnT 在共联体的什么位置？

5. 如果玉米 *Ds* 转座因子采用 Tn3 的转座方式进行转座，看到斑点籽粒的频率还有那么高吗？为什么？

6. 假定有两个无细胞转座系统，各自含有 Tn3 和 Ty 转座必需的全部酶类。下列不同的抑制剂对这两个系统会有什么影响？为什么？

a. 双链 DNA 复制抑制剂；

b. 转录抑制剂；

c. 反转录抑制剂；

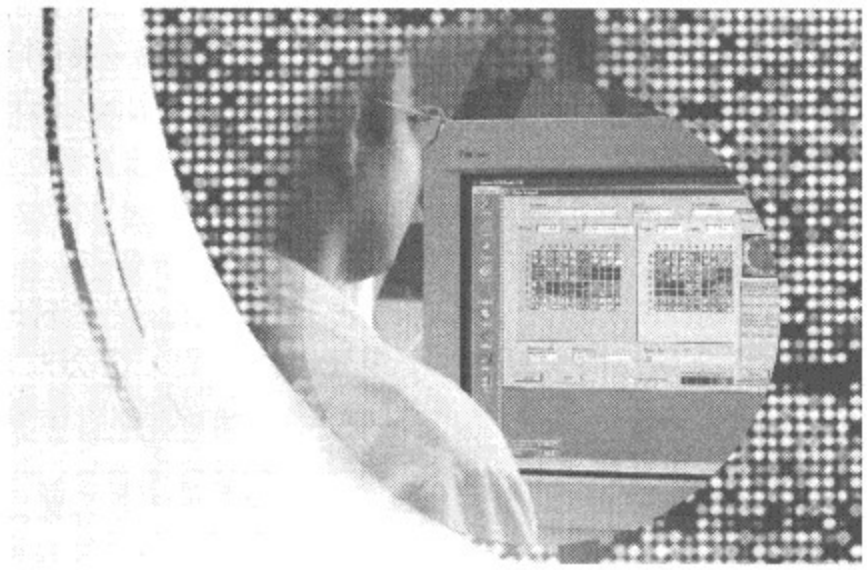
d. 翻译抑制剂。

7. 已鉴定出一种叫做 Rover 的新型转座子，设计一个确定 Rover 是以反转录转座子机制转座还是以 Tn3 所用的复制式转座机制转座的实验，并给出两种情形下分别可能出现的结果。

翻译　陈　璇　张绍鹏

校译　高友军　郑用琏

第 24 章　基因组学、蛋白质组学和生物信息学



DNA 微阵列可同时检测成千上万个基因的表达。图为技术人员正在分析特异基因的表达。（*Source*: *Copyright © IncyteGenomics.*）

在本书的大部分内容里，我们在探讨基因的功能时常常以单个基因为例。然而，随着更为高效且相对不昂贵测序方法的出现，分子生物学家已经能够获得整个**基因组**（**genome**）的碱基序列，并由此产生了一个新的分支学科——**基因组学**（**genomics**），即研究整个基因组的结构和功能的学科。分子生物学家探寻的另一个很类似，但更为复杂的问题是关于生物体的蛋白质组（proteome），即生物体一生所表达的所有蛋白质的特性和活性，故将该领域称之为蛋白质组学（proteomics）。

在本章，我们将讨论科学家用于大规模 DNA 测序和基因表达研究的技术，还要探讨用于蛋白质组学研究的方法，以及从基因组学和蛋白质组学研究中所获知识的应用。

最后，我们将介绍**生物信息学**（bioinformatics），一门管理和应用来自基因组学、蛋白质组学和其他大规模生物学研究所获得的海量数据的学科。

24.1　图位克隆：基因组学介绍

在探讨基因组学研究技术之前，我们先了解一下基因组信息学的一个重要应用：**图位克隆技术**（positional cloning）。图位克隆是一种发现遗传性状相关基因的方法，常用于鉴定人类遗传疾病相关的基因。我们先来回顾一个在基因组时代之前完成的图位克隆实例——对引起人类亨廷顿病（Huntington disease，HD）异常基因的发现。我们会看到，大部分研究都用在缩小查找突变基因的范围上，目的是避免不必要的大规模 DNA 测序。现在这已不再是问题了，因为全基因组测序已经完成了。然而，有理由以本例作为基因组学的引言，以阐明图位克隆的原理，该方法仍然是基因组信息的主要应用，可以说明在缺乏基因组信息的前提下开展图位克隆是多么困难，这也是一个值得传颂的伟大事迹。

图位克隆的传统手段

遗传学家在寻找与控制人类遗传疾病的基因时，常常面临这样一个问题：在不知道突变蛋白特征的情况下寻找未知功能的基因。所以他们必须在人类遗传图谱上确定该基因的位置，这一过程就是图位克隆。

图位克隆策略源于研究家庭或家族性疾病，目的是寻找一个或多个与“疾病基因”（该基因发生突变可导致特定疾病的发生）紧密连锁的标记。通常这些标记并不是基因而是 DNA 片段，其限制酶酶切图谱或其他物理特性在不同个体间有差异。

由于标记的位置是已知的，因此疾病基因可锁定在基因组中一个相对较小的区域。然而，这个“相对较小”的区域常包含约 100 万个碱基对，因此工作并没有结束。下一步就是在这百万碱基中发现可能引起突变疾病的基因位点。传统上，有几种方法用于这种研究，这里介绍两种：一种是外显子捕获法（exon trap）寻找外显子；另一种是对基因的 CpG 岛

进行定位。关于这些方法的应用将在下一节讨论。

限制性片段长度多态性 在20世纪后期，仅有少数人类基因被定位，因此，想定位一个邻近已知基因的新基因都很难。另一种方法则不着眼于发现与已知基因连锁的新基因，而是将新基因与可能不含任何基因的一段“无名”DNA片段建立连锁，通过限制酶酶切图谱识别这些DNA片段。

由于每个人在遗传上都不同于其他人，他们的DNA序列也会有所不同，因而这些DNA的限制酶酶切图谱也会有所不同。以识别AAGCTT的*Hind* III为例，假设在某一染色体的某一区域（如图24.1），一个个体有3个*Hind* III识别位点，分别相距4kb和2kb。另一个个体可能缺少中间的一个位点而保留了另外两个位点。用*Hind* III酶切前者的DNA将产生4kb和2kb的两个DNA片段，酶切后者的DNA仅产生一个6kb的DNA片段，这就是**限制性片段长度多态性**（restriction fragment length polymorphism，**RFLP**）。多态性是指一个遗传位点有不同的碱基组成形式，或不同等位基因（第1章），因此RFLP这一术语的简单含义是指任何两个个体的DNA用同一种限制酶切割后可能产生不同长度的片段。缩略语RFLP的发音通常为“rifflip”。

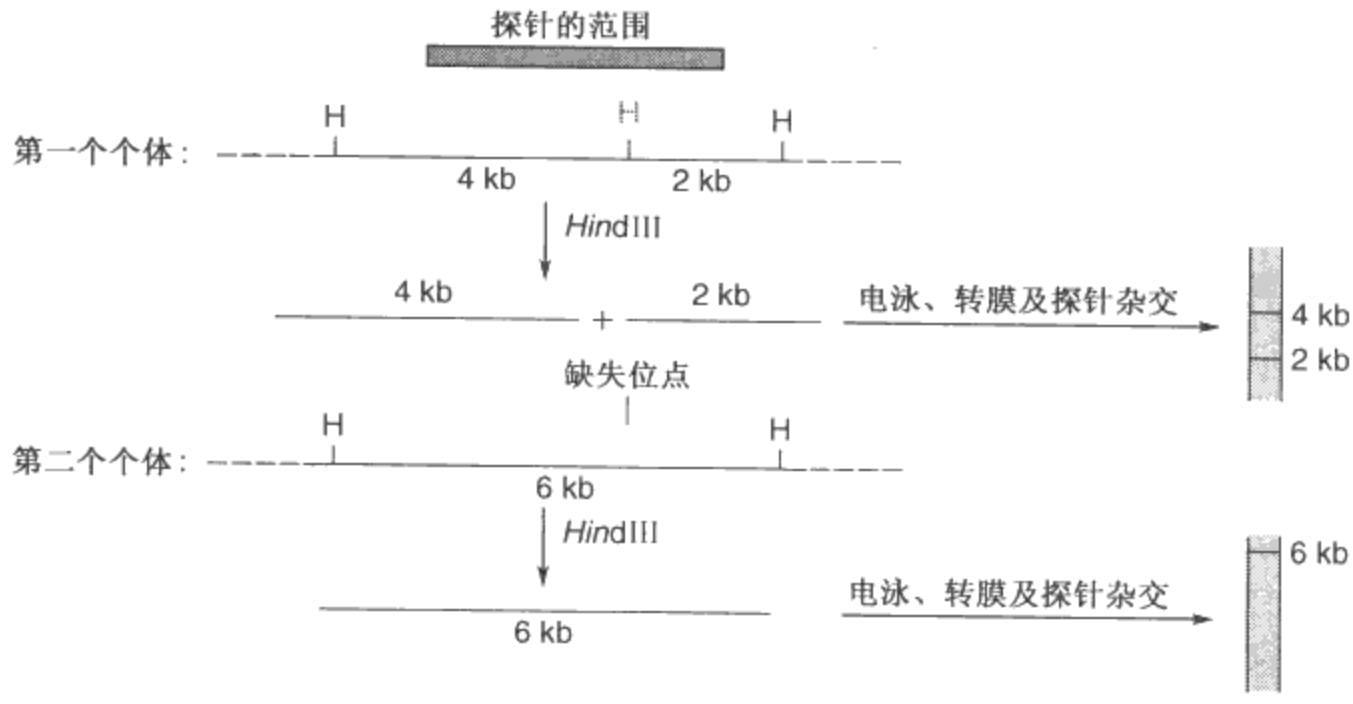


图24.1 检测RFLP。两个个体DNA在*Hind* III限制酶酶切位点（红色）上的多态性。第一个个体含有该位点，用*Hind* III酶切后产生两个分别为2 kb和4kb的片段，能与探针杂交，探针的长度显示在图上部。第二个个体缺乏该位点，用*Hind* III酶切后只产生一个6kb片段，也能与探针杂交。对这些片段进行电泳、转膜后与探针杂交、自显影，结果显示在右边。片段两端的虚线部分因不能与探针杂交，故在图中看不到。

如何着手寻找一个RFLP片段呢？显然我们不能一次对整个人类基因组进行分析。一个典型的限制酶在人类基因组中有几十万个酶切位点，因此一次用一种限制酶对整个基因组切割时，会产生几十万个DNA片段。没有人能对如此庞大复杂的DNA片段进行整理归类以区分不同个体间的细微差异。

所幸的是有一种简便方法。用不同探针进行Southern印迹（第5章）可将总基因组中的一小部凸显出来，从而容易发现个体间的差异。但还存在一个难题，因为每种探针只能与人类基因组的一小部分杂交，所以通过一种探针发现与目的基因连锁的RFLP片段的可能性很小。可能得从几千种探针中筛选并最终发现一个适合的探针。这项工作极为艰辛，但至少是个起点，而且这种方法在发现几种遗传病基因过程中已发挥了关键作用。

外显子捕获 一旦某个基因被锁定在一段几十万碱基的区域内，如何将其从这段DNA中挑选出来呢？如果这段DNA尚未被测序，可以对其测序并寻找可读框（ORF）。可读框被翻译时，在相对较长的区域内不会遇到终止密码子。但寻找可读框还是件很费力的事，有几种更为高效的方法，其中包括Alan Buckler发明的称之为外显子扩增（exon amplification）或**外显子捕获（exon trapping）**的方法。图24.2说明了如何利用外显子捕获法筛选目的基因。首先，我们需要了解Buckler为此方法设计的质粒载体pSPL1，该载体包含一个由SV40早期启动子控制的嵌合基因，此基因源于兔β-球蛋白基因，并以含自身5′-和3′-剪接位点在内的人免疫缺陷病毒（HIV）的一个内含子替换β-球蛋白基因的第二个内含子。将人类基因组DNA片段插入该质粒内含子的一个限制性位点，并将重组载体转入猴COS-7细胞，该细胞能在SV40启动子作用下转录插入基因。如果插进内含子的基因组DNA片段是个完整的外显子并带有自身的5′-和3′-剪接位点，

这个外显子将成为 COS 细胞中加工转录物的一部分。继而在这些 COS 中纯化 RNA，并将其反转录为 cDNA，以该 cDNA 为模板用插入位点两侧区域特异性引物进行 PCR 扩增。这样在引物结合位点之间插入的任何外显子都会得到扩增。最后，克隆获得 PCR 产物，这就是外显子本身。而外显子之外的任何其他 DNA 片段插入该质粒内含子时，由于没有剪接信号，插入片段转录后将与两侧的内含子一道被剪除而丢失。

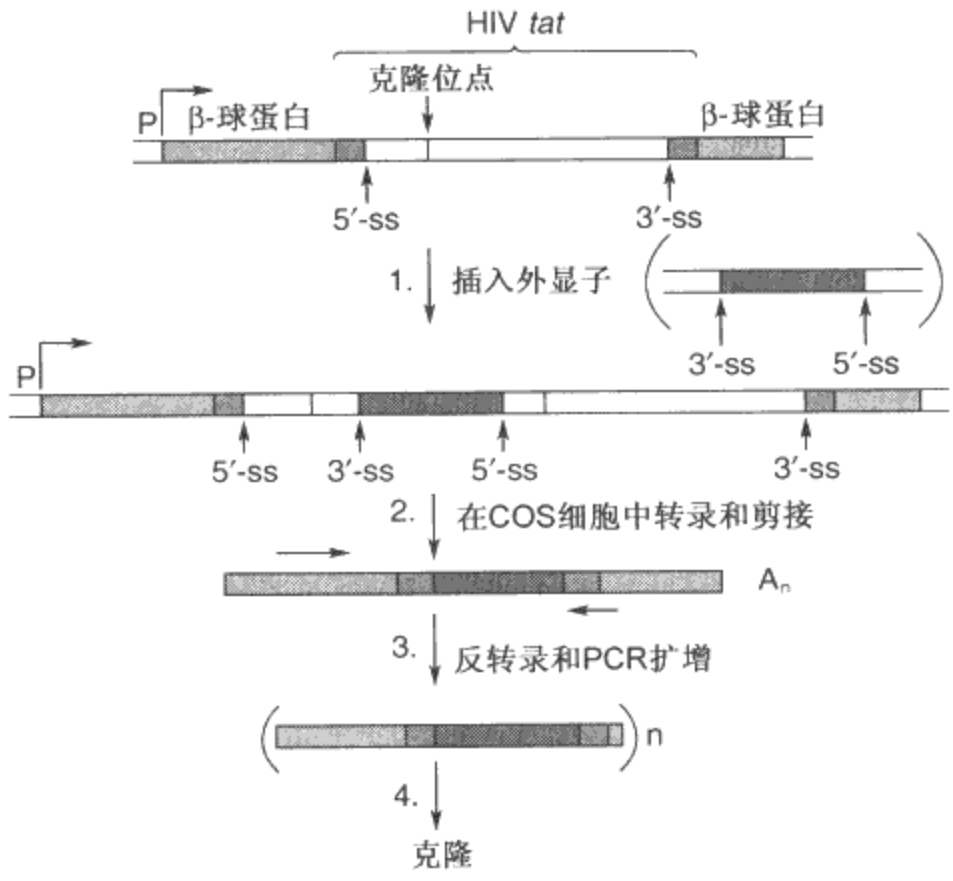


图 24.2　外显子捕获法。以克隆载体 pSPL1 为例，这里显示其简化形式。此载体有一个 SV40 启动子（P），其下游是一个 β-球蛋白基因（橘黄色），其中插入了含有两个外显子（蓝色）和一个内含子（黄色）（位于两个外显子中间）的 HIV *tat* 基因。内含子-外显子间包含 5′端和 3′端剪接位点（ss）。*tat* 基因的内含子包含一个克隆位点，任何 DNA 片段都可插入。步骤 1：外显子（红色）已插入，两侧是自身部分内含子和自身 5′剪接位点和 3′剪接位点。步骤 2：将构建的载体转染 COS 细胞，COS 细胞能转录插入片段，并对转录物进行剪接。注意插入的外源外显子（红色）由于带有自身 5′剪接位点和 3′剪接位点，因而在剪接过的转录物中保留下来。最后（步骤 3 和 4），将转录物进行反转录和 PCR 扩增，引物为箭头所示。获得含有外来外显子的很多拷贝，可用于克隆和测序。注意，非外显子没有剪接位点，与内含子一起剪除，不会留到步骤 3 中，因此无须浪费时间研究。

CpG 岛（CpG island）　另一个发现基因的技术是基于人类活跃基因调控区多含有未甲基化的 CpG 序列，而无活性区的 CpG 序列几乎总是甲基化的。此外，许多甲基化 CpG 位点由于**CpG 抑制（CpG suppression）**现象而在进化过程中丢失。CpG 抑制是指甲基化 CpG 位点中的甲基化脱氧胞嘧啶（methylC）能自发地脱氨基转为甲基化尿嘧啶（methylU），而 U 与 T 是对等的，因此，一旦 methylC 脱氨基即可变为 T。如果这种转变不能被立即识别和修复，在下一轮的 DNA 复制中 T 将与 A 配对，而且这一突变将长期保留。相比之下，未甲基化的 CpG 序列脱氨基产生的 U 通常立即被识别并被尿嘧啶-*N*-糖苷酶（见第 20 章）切除，然后由正常 C 替代，所以基因组中未甲基化的 CpG 序列被保留下来。

另外，限制性内切核酸酶 *Hpa* II 仅在第二个 C 未甲基化时才切割 CCGG 位点。换句话说，它切割含未甲基化 CpG 的活跃基因的 CCGG 位点，保留含甲基化 CCGG 的无活性序列。这样，遗传学家就能从不被 *Hpa* II 切割的 DNA 序列“大海”中扫描大段 DNA 区域，寻找 *Hpa* II 切割位点的“岛屿”。这样的位点被称为 **CpG 岛（CpG island）**，或 **HTF 岛（HTF island）**，因其产生 *Hpa* II 小片段。

小结　图位克隆从基因定位（第 1 章）开始将目的基因锁定到一个相当小的 DNA 区域，定位确定依赖于一套与目的基因所在位置连锁的基因组路标（landmarker）。基因组路标有时是基因，但更多时候是 RFLP 位点，即导致个体间特定限制酶片段长度差异的位点。有几种方法可用于从一大段未测序 DNA 中鉴定基因，其中一种是借助特定的载体克隆外显子的外显子捕获法，还有一种是借助甲基化敏感的限制酶寻找 CpG 岛，即含有未甲基化 CpG 序列的 DNA 区。

鉴定与人类疾病相关的突变基因

我们以精确定位亨廷顿病基因这一经典图位克隆实例来阐明本节内容。

亨廷顿病（Huntington disease，HD）是一种渐进性神经紊乱症。该病刚开始表现为无法察觉的轻微痉挛和行动笨拙。多年后，症状加重，并伴随感情上的错乱。HD 研究者 Nancy Wexler 这样描述这种恶性疾病的晚期症状：整个身体自发性抖动，躯体扭动和面部扭曲，发展到晚期的患者看起来具有戏剧般的表情。经过 10～20 年后，患者最终死亡。

亨廷顿病由一个显性基因控制。因此 HD 患者的孩子患这种疾病的概率为 50%。HD 患者通过不要孩子以避免该病的遗传，只是 HD 患者通常直到出生数年后才出现 HD 最初的症状。

因为人们不知道 *HD* 基因产物的特性，所以遗

传学家不能直接找到这个基因。能想到的最好方法是寻找一个与 *HD* 紧密连锁的基因或标记。Michael Conneally 及同事花了十多年的时间试图找到这样一个连锁基因，但是没有成功。

随后，Wexler、Conneally 和 James Gusell 尝试通过 RFLP 发现与 *HD* 连锁的遗传标记。所幸的是他们找到一个非常大的 HD 家族作为研究对象，生活在委内瑞拉 Maracaibo 湖边的这个大家族的成员从 19 世纪早期就开始罹患 HD。其中，第一位患者是一位女性，她的父亲（可能是个欧洲人）携带有缺陷基因。这个家族的家谱能追溯 7 代以上，而且每一代的人数不同寻常的多：一个家庭有 15～18 个孩子并不少见。

Gusella 及同事知道寻找与 *HD* 连锁的 RFLP 位点可能需测试几百个探针，但是他们真的很幸运，在第一次测试的 12 个探针中，有一个（称为 G8）探针在 Venezuelan 家族中检测到一个与 *HD* 紧密连锁的 RFLP 位点。图 24.3 显示了与探针杂交的 DNA 片段中 *Hind* III 识别位点所在的位置。我们看到总共有 7 个位点，但在这个家族中只发现其中的 5 个。另外两个用星号标记，编号为 1 和 2 的位点可能存在也可能不存在。因此后两个位点具有多态性或变异性。

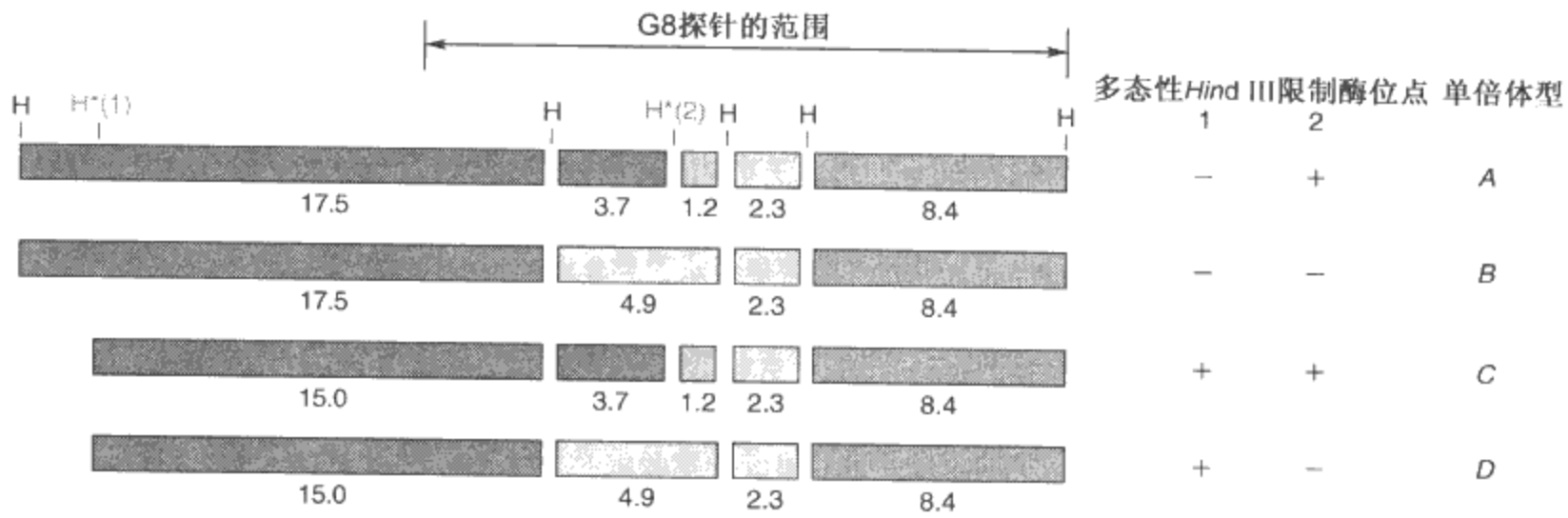


图 24.3　与亨廷顿病基因相关的 RFLP。图中显示了 G8 探针杂交的 DNA 区域内的 *Hind* III 限制酶切位点。对这个家族的研究显示有两个多态性位点，分别用星号数字 1（蓝色）和 2（红色）标记。位点 1 的存在会产生 15kb 和 2.5kb 片段，但是由于 2.5kb 片段不能与 G8 探针杂交，因此无法检测到。位点 1 的缺失仅产生 17.5kb 片段。位点 2 的存在产生 3.7kb 和 1.2kb 片段，而其不存在时则只产生 4.9kb 片段。这两个位点的存在与缺失的 4 种组合产生 4 种单倍体型（A～D），在图的右边列出，旁边是多态性 *Hind* III 位点和 G8 探针在每种单倍体型中检测到的 *Hind* III 片段。例如，单倍体型 A 有位点 2，缺失位点 1，*Hind* III 酶切后产生 17.5kb、3.7kb、1.2kb 片段。2.4kb 和 8.4kb 片段也能被 G8 探针检测到，但是由于它们在 4 种单倍体型都存在，因此不考虑。

让我们看看这两个限制酶酶切位点存在与否是如何产生 RFLP 的。位点 1 缺失，产生一个 17.5kb 的片段。而位点 1 存在时，17.5kb 的片段将被切成 15kb 和 2.5kb 两个片段。由于 2.5kb 片段在 G8 探针杂交区之外，因此放射自显影仅显示出 15kb 片段。位点 2 缺失，产生 4.9kb 片段，而位点 2 存在时，4.9kb 片段将被再分为 3.7kb 和 1.2kb 两个片段。

这两个多态性 *Hind* III 位点有四种可能的**单倍体型**（haplotype）（在单个染色体上的等位基因簇），分别标记为 A～D。

单倍体型	位点 1	位点 2	观察到的片段
A	缺失	存在	17.5kb、3.7kb、1.2kb
B	缺失	缺失	17.5kb、4.9kb
C	存在	存在	15.0kb、3.7kb、1.2kb
D	存在	缺失	15.0kb、4.9kb

术语单倍体型（halotype）是单倍体基因型（haploid genotype）的缩写，强调每个家族成员经遗传而得到两个单倍体型，一个来自父亲，一个来自母亲。例如，一个人从父亲遗传获得 A 单倍体型，从母亲获得 D 单倍体型，因此这个人的基因型（一对单倍体型）为 *AD*。有时不同的基因型无法区分，如一个 *AD* 基因型的人和一个 *BC* 基因型的人有同样的 RFLP 模式，因为所有的 5 个片段在两种情况下均出现。通过检测双亲的基因型可以推测其真实基因型。图 24.4 显示了以放射性 G8 探针对两个家族进行 Southern 印迹分析的放射自显影结果。对 *AC* 基因型而言，17.5kb 和 15kb 片段迁移后非常相近，因此当两者都存在时很难被区分开。然而，仅含 17.5kb 片段的 *AA* 基因型很容易与仅含 15kb 片段的 *CC* 基因型区分。基于 4.9kb 片段的存在，可以判断在第一个家族中存在 B 单倍体型。

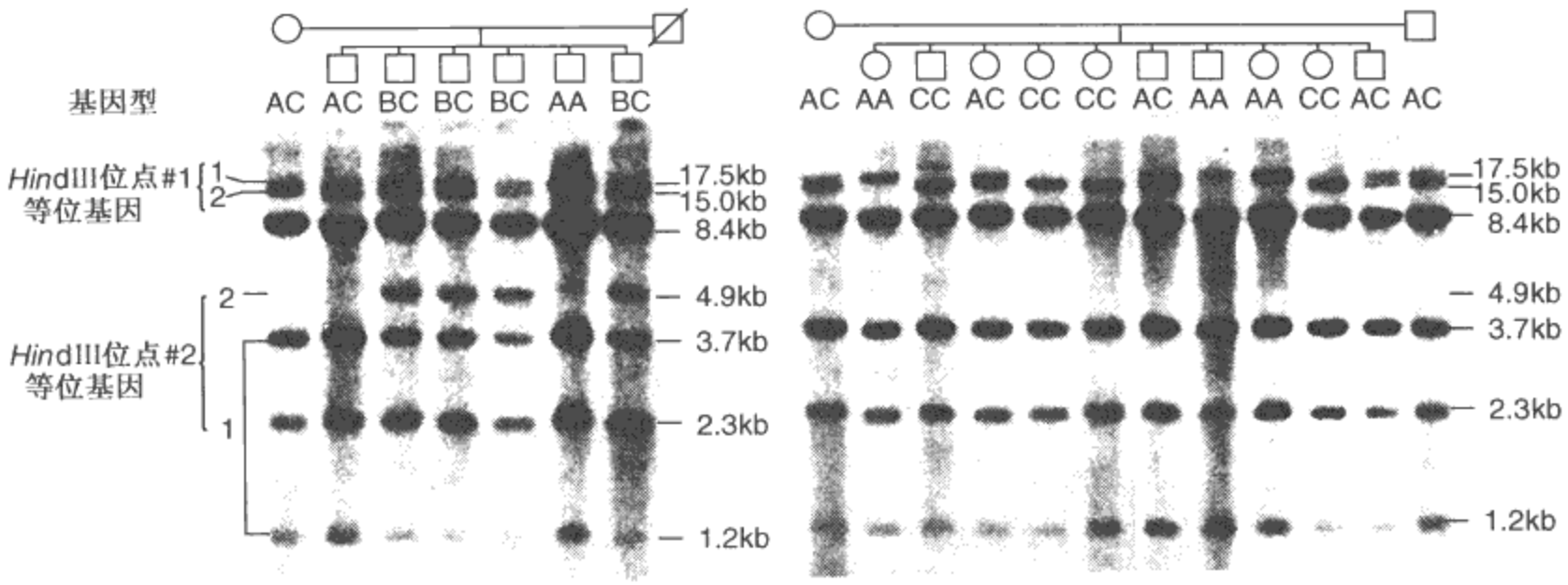


图 24.4　Southern 印迹检测两个家族成员的 *Hind* III 限制酶酶切片段（与 G8 探针杂交）。右边列出放射自显影条带的大小。顶端列出 4 个亲代中的 3 个及其所有孩子的基因型。由于其中一对父母之一死亡，其基因型无法确定。（*Source*：Gusella，J. F.，N. S. Wexler，P. M. Conneally，S. L. Naylor，M. A. Anderson，R. E. Tauzi，et al.，A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* 306：236. Copyright © 1983 Macmillan Magazines Limited.）

到底哪个单倍体型与 Venezuelan 家族的 *HD* 相关呢？图 24.5 能够证明是 *C*。几乎所有携带这种单倍体型的个体都患有这种疾病，即使还没有发病，也几乎可以肯定他以后会患这种病。根据图 24.5 也能断定缺乏 *C* 单倍体型的个体不会患病。这一方法能非常准确地预测该家族成员是否携带 *HD* 相关基因。对一个美国家族进行的类似研究表明，A 单倍体型与家族的 *HD* 连锁。因此，每一家族与 *HD* 关联的单倍体型不同，但在一个家族内，这个 RFLP 位点与 *HD* 的连锁非常紧密，所以这些位点间的重组非常罕见。到此我们可以理解 RFLP 位点可作为图谱定位的遗传标记，就好像它是一个基因。

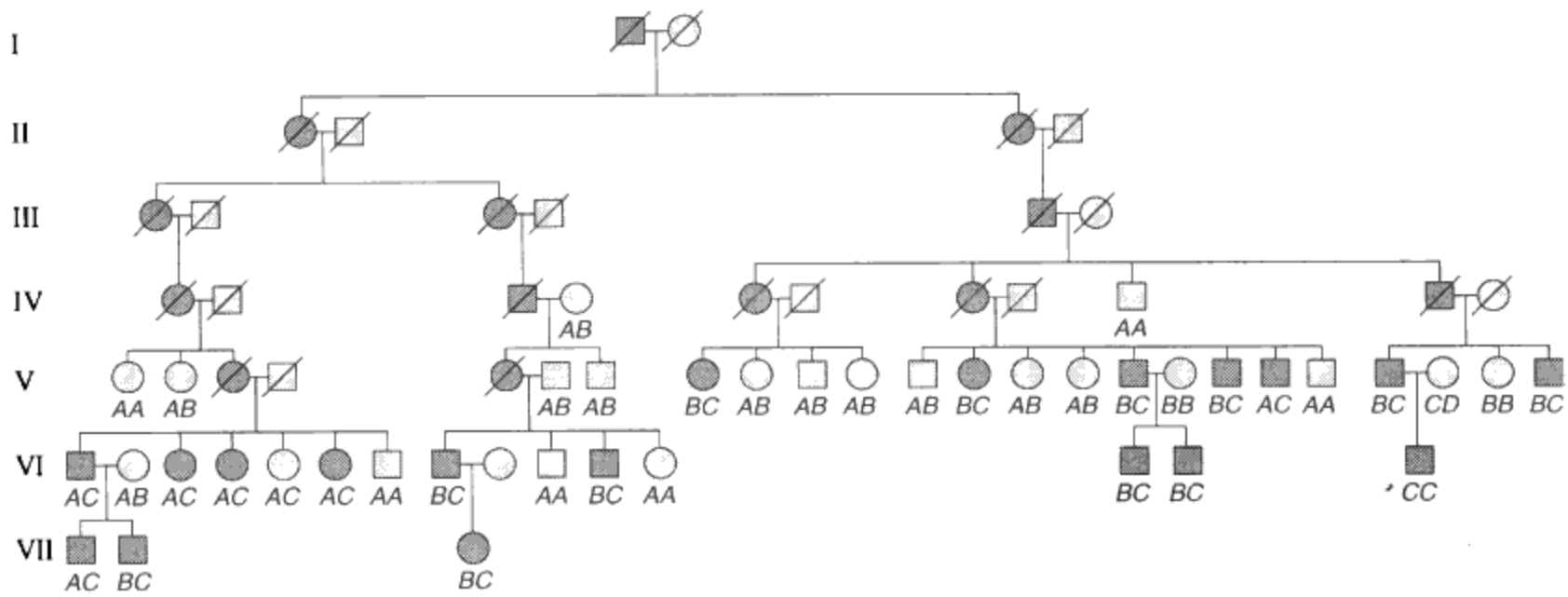


图 24.5　Venezuelan 大家族 HD 病的家系图谱分析。家庭成员中被确定患有此病的用紫色标记。应该注意的是，很多带有 C 单倍体型的成员已经患上此病，未患病的则缺乏 C 单倍体型。这样，C 单倍体型与 HD 病有很高的相关，相应的 RFLP 与 *HD* 基因紧密连锁。

HD 与特定 DNA 区（该区能与 G8 探针杂交）连锁的发现也为 Gusella 及同事将 *HD* 定位到人 4 号染色体上提供了可能，为实现这一目的他们建立了人鼠杂交细胞系，且每个系只含少量的人染色体，将这些细胞系的 DNA 与放射性 G8 探针杂交，发现只有含 4 号染色体的细胞系能与之杂交，与其他染色体存在与否无关。因此可以确认人类 4 号染色体携带 *HD* 基因。

到此为止，*HD* 图谱定位研究组的工作一直都很顺利。而随后将 *HD* 基因定位到 4 号染色体远末端却费了很大的周折。由于该染色体的末端是遗传信息贫乏的重复序列，给研究带来了很多困难。最

后，在徘徊于被 Gusella 称之为遗传“废品站-重复序列”数年之后，Gusella 及研究组把注意力转向更有希望的区域。图谱定位工作提示 *HD* 基因并不存在于 4 号染色体的最末端，而是在距末端数百万碱基的一个 2.2Mb 的区域内。就筛查基因而言，除非你已经知道这段 DNA 的序列，否则超过 2Mb 碱基的 DNA 是一个极为庞大的 DNA，因此 Gusella 决定将研究集中于一个在约 1/3 HD 患者（这些患者似乎有共同的祖先）中均很保守的 500kb 区域。

一个 500kb 的人类基因组区段平均包含约 5 个基因，为了找到它们，Gusella 及同事用外显子捕获法确定了少数几个外显子克隆。继而以这些外显子为探针检测 cDNA 文库，以确定靶 DNA 区转录产物（即 mRNA）反转录的 DNA 拷贝。其中的一个称为 IT15（interesting transcript number 15）的克隆与 cDNA 杂交，并确定该 cDNA 为一较大（包含 10 366个核苷酸）的转录物，编码一个较大（3144 个氨基酸）的蛋白质。根据此 cDNA 序列推测的蛋白质与任何已知蛋白质无相似性，因此没有任何证据说明它就是 *HD* 基因。然而这个基因中重复的 23 个 CAG 三联体拷贝（其中一个是 CAA），即编码一段 23 个谷氨酰胺的肽链，确实很令人好奇。

这真的就是 *HD* 吗？Gusella 小组通过比较 75 个 HD 家族中患 HD 和不患 HD 个体间该基因的差异，结果显示此基因正是 *HD*。在所有未患病个体中，CAG 重复数为 11～34，且 98%的未患病者有 24 或更少的重复数。在所有患病个体中，CAG 重复数至少 42 个，甚至高至 100。因此，通过查看该基因中 CAG 的重复数就可预测一个人将来是否会患 HD。

而且，HD 的严重性或发病年龄与 CAG 的拷贝数有一定程度的相关。患者群中重复数低（现在知道是 36～40）的人通常在症状出现前都能很好地生存到成年，而重复数高的人倾向于在儿童期即出现症状。极端的例子是 CAG 重复数最高（约 100）的个体竟然早在两岁时就开始出现症状。

最后，有两个患者出现症状时其父母还没有出现症状。在这两个病例中，患者的 CAG 重复数增加了，而其父母的 CAG 重复数并没有增加。新的突变（增加 CAG 重复数）虽然在 HD 中很少发生，但还是引发了上述两个病例，这是显而易见的。

另一种证明该基因确实是 *HD* 的方法是对其进行人为突变，并证明这些突变有神经病学效应。很显然，如果这个基因就是 *HD*，不可能在人身上做实验，但对小鼠是可行的。所幸的是 *HD* 在包括小鼠在内的很多物种中是保守的，小鼠中这个基因是 *Hdb*。1995 年，Michael Hayden 小组通过靶向破坏 *Hdb* 的外显子 5 构建基因敲除鼠（第 5 章）。结果纯合的基因敲除鼠死在子宫中。而杂合的基因敲除鼠能够存活，却显示出与神经细胞功能降低相应的智力降低。这一结果强调了 *Hdb*（并推及 *HD*）在大脑中起重要作用，这正是我们对 *HD* 基因所预测的。

怎样将这一新发现应用在实际工作呢？一个显而易见的方法是通过精确的遗传学筛查检测可能患该病的人。实际上，通过计算 CAG 重复数，甚至能预测出发病年龄。然而，这种信息有利也有弊，因为可能造成患者心理上摧毁性的打击。我们真正需要的当然是使患者恢复健康，但那可能是一个漫长的过程。

为找到 HD 的有效治疗方法，至少需要回答以下两个问题。第一，*HD* 产物［现称为 huntingtin 蛋白（Htt）］的正常功能是什么？为了回答这一问题，遗传学家构建了携有增加了多聚谷氨酰胺的 *HD* 基因的转基因小鼠，这些小鼠表现出和 HD 患者一样的神经传递素受体变异。而且，微阵列分析表明转 *HD* 基因小鼠中编码神经传递素受体的基因表达发生了改变，胞内信号发送系统也随之发生了改变。很明显，对神经细胞中神经传递素受体和信号发送系统的破坏会给神经功能带来极大的负面影响。

有意思的是转 *HD* 基因小鼠体内表达水平发生改变的基因也包括一些 Sp1 激活因子的靶点。更有意思的是发现 Sp1 转录激活域富含谷氨酰胺，该区与 $TAF_{II}130$（人类对应果蝇 $TAF_{II}110$ 的同源物，第 11 章）的谷氨酰胺富含区相互作用激活人类基因。一连串的推理提示以下假说：携有过量谷氨酰胺的突变亨廷顿蛋白可能与 Sp1 强烈互作而阻止了 Sp1 与 $TAF_{II}130$ 的结合，继而阻止了 Sp1 激活编码神经传递所需蛋白质相应基因的表达，导致神经功能的衰退和 HD 症状出现。

2002 年，Dmitri Krainc 及同事报道了支持这一假设的证据。首先，他们利用酵母双杂交系统分析证明 Sp1 和 $TAF_{II}130$ 都能与正常亨廷顿蛋白相互作用，且 Sp1（但不是 $TAF_{II}130$）能与突变亨廷顿蛋白发生更强的互作。其次，在培养的转 *HD* 基因鼠细胞中转入并共表达 Sp1 和 $TAF_{II}130$ 基因，结果阻止了突变 *HD* 对神经传递素受体基因（多巴胺 D2 受体基因）转录的抑制。这个结果被认为是转入的 *Sp1* 基因表达了过量 Sp1 所引起，因为这些 Sp1 会占用突变的亨廷顿蛋白。最后，一个证据是 HD 患者死后的脑组织中可溶性的突变亨廷顿蛋白阻断了 Sp1 与其 DNA 靶点结合的能力。

基因组数据的优点　在数年的图位克隆研究中，主要的时间都耗费在对可疑 DNA 区段的测序，并

力图鉴定出此序列中哪个基因是可能的“肇事者”。随着人类基因组计划的完成，图位克隆已经变得比较容易。正如多年来一直从事小鼠基因图位克隆研究的遗传学家 Neal Copeland 所说，“有基因组序列前，我们经历 15 年获得了 10 个可能的癌症相关基因，而一旦有了全基因组序列，只需数月就一次性得到了 130 个基因”。当然他说的是小鼠基因组，但同样的原理也适用于人类。且小鼠图位克隆研究鉴定出的基因常常也能在人类中引起类似的症状。因此，基因组研究最大的好处是加速了发现人类疾病基因的进程。但不能因此认为图位克隆已经过时，只要我们仍然有兴趣寻找生物体的性状相关基因，图位克隆就将继续发挥重要作用。基因组全序列的获得只是使图位克隆更为容易。

小结 利用 RFLP，遗传学家将亨廷顿病相关基因（*HD*）定位到邻近 4 号染色体末端的一个区域。继而用外显子捕获法确定了该基因。引起这种疾病的突变是基因中 CAG 重复序列从正常的 11～34 个拷贝增加到至少 38 个拷贝。这些额外的 CAG 重复序列导致额外的谷氨酰胺插入亨廷顿蛋白。

知识窗 24.1
遗传筛查中的问题

既然我们已经确定了引起 Huntington 症、胞囊纤维化（cystic fibrosis，CF）以及其他一些严重疾病的基因，如何应用这些知识呢？我们确实能够鉴定出这些遗传病的携带者以及这类疾病的潜在患者，这听起来是个好事，一般而言这的确是好事，但也存在一些伦理上的问题。

来看一个实例，收养所打算安置一个母亲患有 HD 的女婴，欲收养者想知道这个女婴是否遗传了致病基因。然而，被咨询的遗传学家 Michael Conneally 决定不对这个女婴做检测。理由如下：如果检测结果呈阳性且以后被女孩知道，她将知道自己注定要患 HD 症，许多这类疾病的潜在患者无法接受自己携带遗传病基因的事实，因此他们选择不做这项检测。而且在此例中，女婴太小还不能由自己做出选择。

最近，决定是否需要知道自己是不是 *HD* 基因携带者的问题变得更为复杂。现在通过计算 *HD* 基因中 CAG 的重复次数，就可以预测 HD 发病年龄。以前人们可能会侥幸地认为，即使一个人携带了 *HD* 基因，HD 还可能在其很老的时候才发生。现在连这样的希望也因遗传检测破灭了。

另一种进退两难的处境会在如下情形中发生：一女性的父亲是 HD 患者，因此她有患此病的可能，而一旦知道自己遗传了 *HD* 基因，她将无法承受，于是决定不做检测。随后，她怀孕了，和丈夫想要一个没有遗传缺陷的孩子，否则就不要。他们可以通过羊水诊断检测胎儿的基因型，如果结果是阴性的，这对夫妇对孩子的忧虑消除，并且母亲对自己的未来满怀希望，但是阳性结果对她自己及婴儿都是极为恐怖的。

对于常染色体隐性遗传的胞囊纤维化症（CF）而言又会引发其他的一系列问题。胞囊纤维化症是美国儿童中最常见的致命遗传病。1/29 的白种人是 CF 携带者，且超过 1/4000 的白种人儿童为 CF 纯合子并患此症。既然 *CF* 基因已被鉴定，预计数百万人想利用这项检测来帮助他们制定生育计划。然而，1989 年，美国人类遗传学会要求暂停这项普查。

有两个主要阻碍。首先，尽管我们期望 CF 仅由少量几种突变引起，但这些突变的数量在 2001 年已超过 900 个，其中约有 20 个突变是常见的。即使这样，我们也只能发现所有携带者中的 85%～90%。即使检出率达到 90%，还是只能检出高危夫妇中的 81%（0.9×0.9＝0.81）。其余近 20%的夫妇由于漏检可能会生一个患病儿，这样的结果远不理想，检测机构也担心给出的建议不够完善。

即使检测结果非常理想，仍然有一些问题。主要的困难是检测任务很庞大。因为 CF 症很常见，大部分即将成为父母的人可能都需要做这项检测，仅在美国，就有约 1.5 亿人口处于生育年龄，即使可以检测这么多人，还得有人向他们解释检测结果意味着什么，我们远没有这么多专业的遗传学顾问。

一个考虑不充分的计划将带给我们什么？筛查美籍非裔镰状贫血症（sickle cell disease）的经验给我们留下了教训。在20世纪70年代早期，美国许多州颁布法律要求美籍非裔检测是否携带镰状贫血症基因，但是却没有提供恰当的咨询服务，造成令人遗憾的结果。其一是许多仅为镰状贫血症基因的携带者被错误地判定患有此病。还有，建议镰状贫血基因的携带者不要孩子，这在白人中很常见，但却被一对美籍非裔夫妇携带者以种族歧视起诉。而且，某些携带者的隐私权没有得到尊重，导致人寿保险公司拒绝为某些镰状贫血症基因携带者投保。这样的结果使原来旨在提供帮助的计划失去了人们的信任。因此，很少有人做检测，采纳检测结果为其计划生育提供帮助的人就更少了。

尽管存在这些问题，在2001年底时，经改进的*CF*基因检测的准确性已经相当好了。美国妇产医师学院（American College of Obstetricians and Gynecologists，ACOG）提出了两个建议。一是倡导为任何想进行CF检测的怀孕妇女做检测；二是倡导为CF高危的同种族夫妇提供关于该病的信息。他们出版了一本名为《胞囊纤维化携带者检测：由你做主》（*Cystic Fibrosis Carriers Testing*：*The Decision is yours*）的小册子，用通俗易懂的语言和专业词汇对胞囊纤维化症的生物学和遗传学特性、检测方法和检测结果的意义做了解释。医生可以很方便地将这本小册子推荐给被检者。

以上这些例子展示了人们在必须进行遗传检测时所面临的艰难选择，以及这类测试可能给人们带来的心理冲击。而涉及强制检测时又会产生截然不同的问题。例如，保险公司要求遗传检查是否符合伦理？HD或阿尔茨海默（Alzheimer）病检测阳性的人买健康保险时可能会遭到保险公司的拒绝，或者被迫支付过于高昂的保险金。这样的事在镰状贫血症遗传基因携带者中发生过。如果被保险人由于自己所不能控制的基因而负担过重是不公平的，那么在买人寿保险前被要求做基因检测公平吗？要知道人寿保险公司总是在客户投保时要求健康证明，基因检测被简单地认为是健康证明的一部分。

让我们进一步思考这个问题。假设人寿保险公司支付了基因检测费用，并且发现某个人可能会发展为阿尔茨海默病，假如保险公司与其他人共享这一信息——如患者的老板或者将成为他老板的人，这是否符合伦理？这些信息毫无疑问将对当事人长期的健康状况给出一些暗示，这些信息也必定会使当事人从此背负沉重的心理负担，甚至导致其失业。在一些大公司，如果保险业者就是雇主，问题就更麻烦了。

这些棘手的问题没有简单的答案，但是随着越来越多的遗传病检测手段的改进，我们必须面对这些现实的问题。最后，如果遗传学研究确实可以找到有效的治疗手段，这些问题将迎刃而解。

24.2 基因组测序

正如所预期的，第一个完成全基因组测序的基因组一定是非常简单的。*E. coli* ΦX 174噬菌体的基因组就非常简单，Frederick Sanger（DNA测序的双脱氧链末端终止法发明人）于1977年测得这个5375nt的基因组的序列。

从这个序列中可以获得哪些信息呢？首先，可以精确地对所有基因的编码区定位，精确地判断基因间的空间关系及基因间的距离。如何识别编码区呢？编码区包含一个可读框（ORF），即一段相对较长的足以编码一个噬菌体蛋白的区段。而且ORF总是起始于三联密码ATG（偶尔是GTG），对应于翻译起始密码子AUG（或GUG），并于终止密码子（UAG、UAA、UGA）处结束，即细菌或噬菌体的ORF与基因的编码区是一致的。

还可以从该噬菌体的碱基序列推断其所有蛋白质的氨基酸序列，只需要依据遗传密码子表将每个ORF的DNA碱基序列翻译成对应的氨基酸序列即可。这项工作似乎很费力，但个人电脑瞬间即可完成。

Sanger对ΦX174噬菌体ORF进行分析的结果有些出乎预料且令人迷惑，有些噬菌体基因是重叠的。如图24.6a所示，基因*B*的编码区在基因*A*中，基因*E*的编码区在基因*D*中。而基因*D*和*J*的编码区有一个碱基的重叠。两个基因如何占据相同的位置却能编码不同的蛋白质呢？这是因为两个基因以不同的ORF进行翻译（图24.6b）。在这两个ORF内会遇到完全不同的密码子串，两个蛋白质产物也就完全不同了。

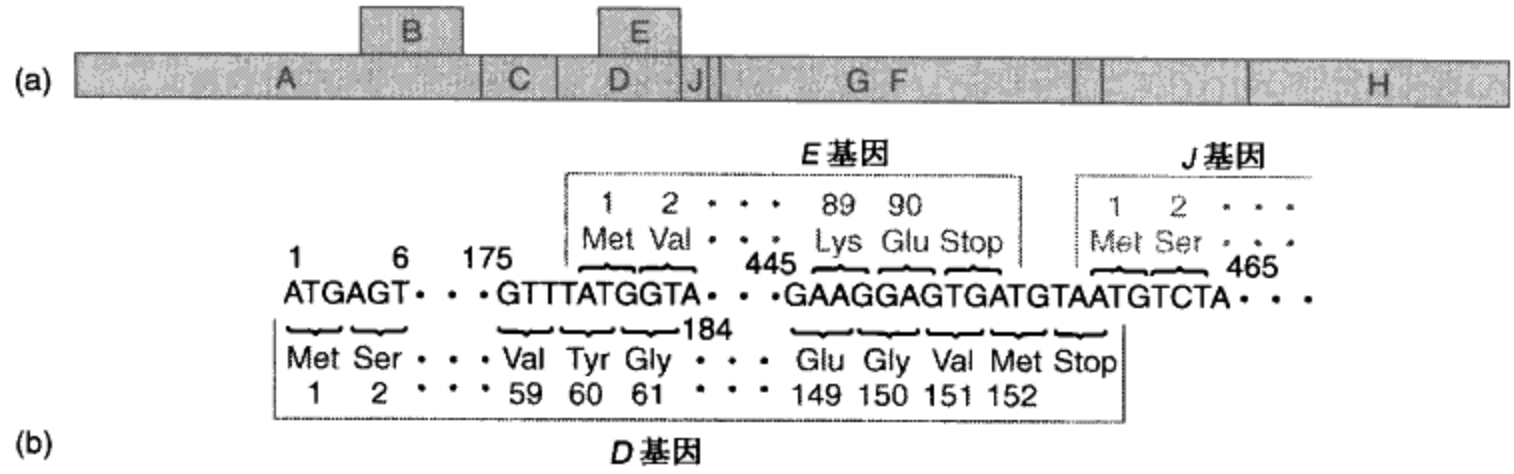


图 24.6　噬菌体 ΦX174 的遗传图谱。（a）每个字母代表一个噬菌体基因。（b）ΦX174 的重叠可读框。基因 *D*（红色）从图中所标的 1 号碱基开始一直到 459 号碱基结束，对应于氨基酸 1～152 位，外加一个终止密码子 TAA。图中的点表示未标出的碱基或氨基酸。图中只显示了非模板链。基因 *E*（蓝色）从 179 位碱基开始到 454 位结束，对应于氨基酸 1～90 位，另加一个终止密码子 TGA。该基因的可读框比基因 *D* 向后移动了一个碱基。基因 *J*（灰色）从 459 号碱基开始，其可读框只在左边与基因 *D* 有一个碱基的重叠。

这的确是一个有趣的发现，随之而来的问题是，这种现象普遍吗？迄今为止，基因重叠现象几乎为病毒所特有。这不奇怪，因为这些拥有小基因组的简单遗传系统，提高遗传物质的使用效率极为重要。加之病毒有很惊人的复制能力，在无数次的传代过程中，进化使其基因组更加完美。

随着自动测序技术的出现，遗传学家在测得全序列基因组的清单中增加了一些更大的基因组。1988 年，D. J. McGeoch 及同事公布了一个重要的人类病毒（单纯疱疹病毒 I）的序列，该病毒的基因组相对较大，为 152 260bp。1995 年 Craig Venter、Hamilton Smith 及同事测定了流感嗜血杆菌（*Haemopbilus influenzae*）和生殖道支原体（*Mycoplasma genitalium*）两种细菌基因组的全部碱基序列。流感嗜血杆菌（Rd 株）基因组全长 1 830 137bp，是第一个被测定的非寄生（free-living）生物基因组全长序列。生殖道支原体基因组仅 580 000bp，是所有已知非寄生生物中最小的基因组，仅含大约 470 个基因。

1996 年 4 月，国际实验室联盟（International Consortium of Laboratory）的领导宣布了另一个基因组测序史上的里程碑：长达120 000 000bp 的酿酒酵母（*Saccbaromyces cerevisiae*）基因组全长测序完成。这是第一个被测定的真核生物全长基因组。1996 年晚些时候，生命的第三种形式古菌（archaea）的首个生物体［詹氏甲烷球菌（*Methanococcus jannaschii*）］的基因组全长被测定。

接着，1997 年，期待已久的 *E. coli* 基因组全长 46 000 000bp 序列被报道。虽然长度仅是酵母基因组的大约 1/3，但是 *E. coli* 作为遗传学研究工具的重要性使得这项工作同样具有里程碑意义。

1998 年，第一个动物基因组，秀丽隐杆线虫（*Caenorbabditis elegans*）基因组被报道。2000 年，第一个植物基因组，十字花科植物拟南芥（*Arabidopsis thaliana*）的基因组测序完成。由于秀丽隐杆线虫和拟南芥的基因组小、生活周期短、遗传实验易于操作而被选为**模式生物**（model organism）用于研究。秀丽隐杆线虫的另一个有利条件是其只有不到 1000 个细胞，而且是透明的，因此它每一个细胞的发育都能够被可视化跟踪。另外两个有名的模式生物是果蝇（*Drosopbila melanogaster*）和小家鼠（*Mus musculus*），这两种生物的基因组分别于 2000 年和 2002 年报道。也是在 2000 年，热切期盼的人类基因组序列草图公布。2001 年，人类基因组“工作草图”发表。

2002 年，一些重要的基因组（至少是以基因组草图的形式）陆续被报道。其中包括引起疟疾的单细胞寄生虫恶性疟原虫（*Plasmodium falciparum*）及其主要携带者疟蚊（*Anopbeles gambiae*）的基因组均见报道。这些基因组的揭晓无疑对我们设计更好的方法来控制疟疾有很大帮助。两种常见水稻（*Oryza sativa*）品种的基因组草图也是在 2002 年公布的，这是第一个完成测序的谷类植物基因组，此项工作在人类营养方面具有深远的意义，因为谷类尤其是稻米是世界大部分人口的主食。

2002 年，另外两种脊椎动物，东方红鳍鲀（*Fugu rubripes*）和小家鼠的基因组序列发表。将这些序列与人类基因组序列比较，可进一步认识脊椎动物的进化历程。另外，玻璃海鞘（*Ciona intestinalis*）基因组序列的揭示也有助于进化研究。玻璃海鞘的成体为固着生物，常吸附于岩石和港口桩基上。虽然海鞘与脊椎动物类似的方面很少，但是其幼虫形似蝌蚪，由软骨组织构成的脊柱结构有些类似于脊骨。因此，海鞘是脊索动物，在分类上与脊

椎动物同属一个门。将海鞘基因组与脊椎动物以及线虫、果蝇等非脊椎动物的基因组进行比较，将使我们对脊椎动物的进化有了更深刻的理解。

大部分分子进化研究是基于对不同生物基因组部分序列的比对。指导原则是任何两种生物基因组序列间的差异与这两个生物体间的进化距离存在相关性。生物体基因组分支相对较近的（如人和鼠）比分支较远的（如海鞘和人）亲缘关系近。一般而言，这是对的，但各种生物体基因组学的研究发现了一些预料之外的特性。例如，人类基因组的进化速率在不同区域并非一致，改变速率相对较快的区域点缀着一些变化速率始终相对较慢的区域，找出这种差别的真实原因引起了人们的强烈兴趣。

到目前为止，通过基因组研究，发现基因组的大小与生物的复杂性相关（此外，在第 2 章讨论 C 值悖论时，发现有很多例外现象）。根据这个原则，原核生物基因组比真核生物基因组要小得多。有趣的是基因组大小与生物复杂度之间的相关性存在相反的情况。例如，目前已经测序的最小真核生物基因组是兔脑炎原虫（*Encephalitozoon cuniculi*）——一种人类及其他哺乳动物的胞内寄生虫，其基因组仅 2.9Mb，只有1997个可能编码蛋白质的 ORF（寄生生活方式使生物体利用很少的基因就能生存，因为可依靠宿主提供所需）。与之相比，在 60 种左右（至 2002 年中期）已测序的细菌基因组中，最大的是固氮菌的百脉根瘤菌（*Mesorbizobium loti*）基因组，全长为 7.6Mb，约含 8000 个基因，甚至比芽殖酵母（budding yeast）细胞的基因组还大。

2003 年 4 月 14 号，国际人类基因组测序联盟（International Human Genome Sequencing Consortium）宣布人类基因组测序比预定计划提前两年完成。更确切地说是完成了测序任务的 99%，这可能是 2003 年科技界的重要成果，测序错误率仅为 1/100 000，所有的序列都按正确的顺序排列出来。尽管仍留有极为复杂的几百个集中在重复区和着丝粒区的缺口需要补齐，但这已经比两年前公布的草图有了很大的改进。

根据基因组在线数据库（www. genomesonline. org）的消息，截至 2006 年 10 月 30 日，已完成 453 个基因组全序列的测定，另有 2202 个正在进行。表 24.1 按时间顺序呈现了基因组测序史上的重要成就，在接下来的几节，我们将讨论从这些序列中获得的知识。

表 24.1　基因组测序的大事记

基因组（重要性）	大小/bp	年份
ϕX174 噬菌体（第一例基因组）	5 375	1977
λ 噬菌体（大-DNA 噬菌体）	48 513	1983
单纯疱疹病毒 I 型（大-DNA 真核病毒）	152 260	1988
流感嗜血杆菌（细菌，第一例生物体）	1 830 000	1995
生殖道支原体（细菌，最小的基因组）	580 000	1995
酿酒酵母（酵母，第一例真核生物）	12 068 000	1996
詹氏甲烷球菌（第一个古菌）	1 660 000	1996
大肠杆菌（研究得最清楚的细菌）	4 639 221	1997
秀丽隐杆线虫（第一个动物，圆形蠕虫）	97 000 000	1998
人类 22 号染色体（第一个人染色体）	53 000 000	1999
拟南芥（第一个植物，十字花科）	120 000 000	2000
果蝇（理想的遗传模型）	180 000 000	2000
人类［基因组学“圣杯”（holy grail）工作草图］	3 200 000 000	2001
恶性疟原虫（疟疾寄生虫）	23 000 000	2002
疟蚊（疟疾的主要携带者）	278 000 000	2002
东方红鳍鲀	365 000 000	2002
小家鼠	2 500 000 000	2002
玻璃海鞘（海鞘，原始脊索动物）	117 000 000	2002
家犬（*Canis lupus familiaris*）（狗，工作图）	≈2 400 000 000	2003
鸡（*Gallus gallus*）（鸡，第一个农场动物）	1 050 000 000	2004
人类（测序完成）	3 200 000 000	2004
水稻（*Oryza sativa*）（第一个谷物）	489 000 000	2005
黑猩猩（*Pan troglodytes*）（与人类最亲缘的物种）	≈3 000 000 000	2005
3 种锥体虫［美洲锥虫（*Trypanosoma cruzi*）、布氏锥虫（*T. brucei*）、利什曼原虫（*Leishmania*），引起严重人类疾病］	25～55 000 000	2005

小结 病毒以及从噬菌体到微生物、动物和植物的不同生物的碱基序列已经获得。人类基因组草图和完成版的人类基因组也已获得。相近物种和差异较大物种基因组间的比较使物种进化关系清晰地显现出来。

人类基因组计划

1990 年，美国遗传学家决定着手一项艰巨的探索，对整个人类基因组进行作图并最终测序，即人类基因组计划（Human Genome Project）。该计划很快成为一项国际性工程，同时也引起一些争议，因为该计划的最终目标是测得人类每一条染色体的碱基序列，而这一目标的实现要耗费大量的人力和巨额成本。高成本是由于人类基因组多达 30 亿个碱基对。为了使大家对这项工作的艰巨性有所了解，我们不妨假设将 30 亿碱基对写下来，这将写满自然杂志 50 万页。如果不嫌麻烦，按平均每秒阅读 5 个碱基，每天读 8h，需要花费 60 年时间才能将它读完。在 1990 年，按照测定一个碱基花费一美元计，整个测序工程需要花费 30 亿美元，远远超过我们过去在一个生物学研究项目上的投入。最终，更有效的测序方法的出现大大加快了基因组计划的完成，成本仅为最初的 10%，低于原先的估计数。

最初的人类基因组计划是系统而保守的。首先，遗传学家需要绘制基因组的遗传图谱和物理图谱，这些图谱包含一些标记（或称为路标），以便随后测出的序列能按正确顺序拼接。大规模测序只能在遗传图谱与物理图谱已经完成，且图谱中所有位点对应的克隆已经获得的情况下才能进行。最初预期整个测序计划全部完成的时间为 2005 年。

1998 年 5 月，由 Craig Venter 创立的以营利为目的私人公司（Celera 公司）着手对人类基因组和其他生物基因组进行测序，令基因组学界震惊的是该公司宣布将在 2000 年之前完成人类基因组草图。这一时间表已经非常惊人了，而更引人注目的是 Venter 提出的新测序方法。与依赖图谱来整理拼接克隆序列的方法不同，Venter 提出了**鸟枪法**（shotgun sequencing）测定基因组序列，即将整个人类基因组随机打断并克隆，继而对这些克隆进行随机测序，最后利用强有力的计算机程序发现重叠序列并将这些序列拼凑起来。此后不久，人类基因组计划的公共财务主管 Collins 接受 Venter 的挑战并宣布也将在 2000 年之前公布人类基因组草图，并在 2003 年之前完成基于先作图再测序策略的完善的基因组序列图。

这场竞争以各方合作而告终。2000 年 6 月 26 号，Venter、Collins 与克林顿总统及其他政要共同出席了在白宫的东宫举行的仪式，宣布人类基因组草图的完成。以下我们将了解对大基因组进行测序的两种方法，一种是先作图再测序（逐步克隆法，clone by clone），另一种是鸟枪法。先来了解一下用于人类基因组计划这样庞大工程的克隆载体。

用于大规模基因组计划的载体

无论用哪种测序策略，首先得把目的片段克隆到合适的载体上，大片段克隆尤其可贵。这里介绍两种常用载体：酵母人工染色体和细菌人工染色体。首先是酵母人工染色体，它曾用于早期的绘图工作。

酵母人工染色体（yeast artificial chromosome，YAC） 第 4 章曾提到克隆工具的主要问题是不能携带足够大的 DNA 片段，无法用于人类基因组物理图谱的绘制。Cosmid 克隆载体仅容纳约 50kb 的 DNA 片段，这对跨越百万碱基的区域作图是远远不够的。

被称为酵母人工染色体的载体在人类基因组作图方面非常有用，因为可以容下上百万个碱基。含有多于兆碱基（megabase）的 YAC 叫做“兆 YAC”（mega YAC）。YAC 含有左右两个酵母染色体端粒（第 21 章），是保护染色体末端所必需的，还含有一个酵母着丝粒，是分裂的酵母细胞姐妹染色体向两极分离所必需的。着丝粒邻接左端粒，巨大的人类 DNA（或其他任何 DNA）可插入着丝粒和右端粒之间，如图 24.7 所示。大段的插入片段可通过用一种限制酶短时间消化很长的人类 DNA 而制备。带有巨大 DNA 插入片段的 YAC 被导入酵母细胞，在这里像正常酵母染色体一样复制。

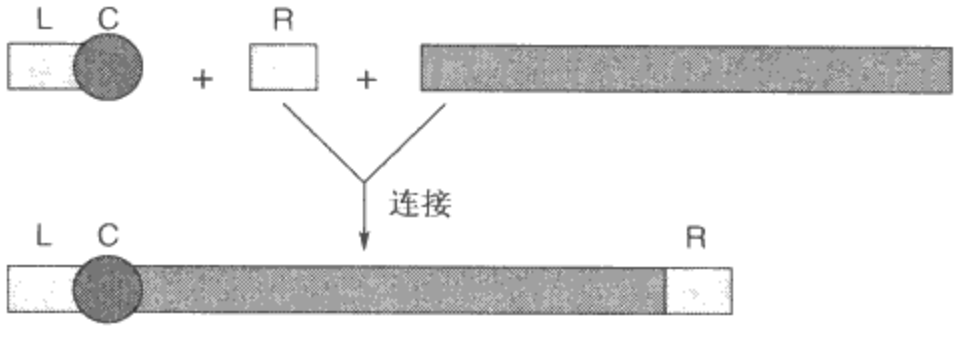


图 24.7 酵母人工染色体的克隆。 左臂含左侧的端粒序列（标为 L 的黄色部分）和着丝粒序列（标为 C 的红色部分）；右臂含右侧的端粒序列（标为 R 的黄色部分）；两臂之间为大的 DNA 片段（蓝色部分），如上百万碱基的人类 DNA。这样形成的 YAC 在酵母细胞中能像正常染色体那样复制。

利用 YAC，遗传学家在人类基因组计划的作图阶段取得了长足的进步，他们绘制出了分辨率为

0.7cM 的全基因组遗传图谱。每厘摩表示在两个遗传标记之间产生 1%重组率的距离，对人类基因组而言，每厘摩平均约为 1Mb 的 DNA 长度。研究者还绘制了两个最小染色体（21 号染色体与 Y 染色体）的较高分辨率的物理图谱。这些图谱非常重要，描绘了 YAC 中克隆的 DNA 重叠片段。在人类基因组被测序前，如果对染色体图谱上某一致病基因感兴趣，我们的任务相对就简单多了，只需找到致病基因两侧的遗传标记，接着找到含有这些遗传标记的一个 YAC 或多个 YAC，最后寻找该基因。

细菌人工染色体（bacterial artificial chromosome，BAC） 尽管 YAC 在人类基因组图谱绘制中取得了成功，它也有很多严重缺陷。例如，效率低（每微克 DNA 获得的克隆不多），从酵母中分离 YAC 较难，稳定性差，存在不同位点 DNA 片段组成的插入片段。细菌人工染色体克服了上述缺点，被应用于人类基因组计划的测序阶段。

BAC 是以大肠杆菌的天然质粒——**F 质粒**（F plasmid）为基础构建的大容量载体系统。该质粒允许细菌细胞间的结合（conjugation）。在有些结合事件中，F 质粒自身被从供体 **F^+ 细胞**转入受体 **F^- 细胞**，使后者转变为 F^+ 细胞，而有时 F 质粒与插入其中的一小段宿主细胞 DNA（插入外源基因的 F 质粒称为 F′质粒）共同转进受体细胞，甚至有时 F′ 质粒插入到宿主染色体中，然后将整条染色体从供体细胞转入受体细胞。由于大肠杆菌染色体大于 400 万 bp，F 质粒显然能容纳大片段 DNA 的插入。实际上，BAC 载体插入片段的长度通常不超过 300 000bp（平均为150 000bp），而且无论在体内还是体外都很稳定。不像线性 YAC 那样易被剪切，环形超螺旋的 BAC 不易被破坏。

图 24.8 显示第一个 BAC 载体图谱，由 Melvin Simon 及同事于 1992 年构建。它有一个复制起点，一个含两个限制酶位点（*Hind* III 和 *Bam*H I）的克隆位点，允许较大 DNA 片段插入，还带有一些基因（*Par* 基因），该载体负责把质粒分配给子细胞，并使每个细胞中质粒的拷贝数维持在 2 个左右，以利于质粒的稳定。此载体还含有氯霉素抗性基因，用以筛选含质粒的细胞。

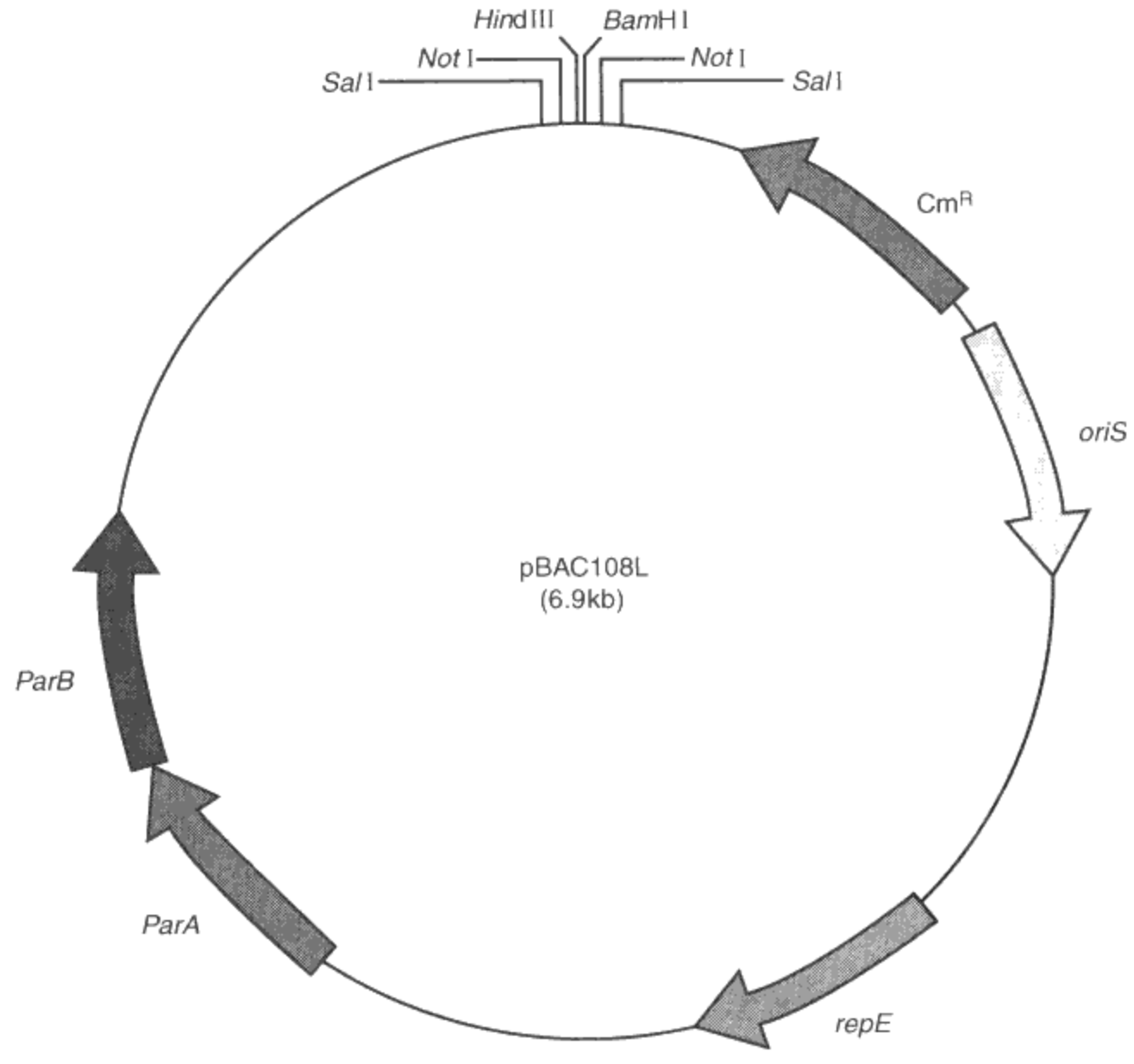


图 24.8 BAC 载体图谱。主要特点是：顶部有两个克隆位点 *Hind* III 和 *Bam*H I；用作筛选工具的氯霉素抗性基因（Cm^R）；复制起点序列（oriS）；调控质粒分配到子代细胞的基因（*ParA* 和 *ParB*）。

小结 两个大容量载体已在人类基因组计划中得到广泛应用。YAC 载体能接受 100 万碱基或更大的 DNA 片段插入，大部分绘图工作主要是利用该载体完成的。BAC 载体能接受约300 000 bp 以内的 DNA 片段插入，大部分测序工作是利用 BAC 实施的。相对 YAC 而言，BAC 更稳定且更易操作。

逐步克隆策略

该策略因系统性强而极具吸引力。首先，通过寻找沿每条染色体均匀分布的标记来绘制全基因组图谱。作图的同时收集各标记相应的克隆群。由于已知各克隆的次序，对每一克隆测序并可将测得的序列置于基因组合适的位置，这种方法称为**逐步克隆测序法**（clone-by-clone strategy）。除用于克隆外，遗传和物理图谱还有另一个重要用途，即为寻找致病基因提供路标。学习本节时，切记这些技术原本是用来定位标记的，这些标记并不是基因，而是一段个体间有差异的 DNA。我们已经提到过一个这样的标记，即限制性片段长度多态性（RFLP）。

可变数串联重复序列（variable number of tandem repeat，VNTR） 某种 RFLP 的多态性程度越高，其用途越大。例如，在 100 人中，若只有一人是一种 RFLP 形式（如图 24.1 中的 6kb 片段），其余 99 人是别的形式（4kb 和 2kb 片段），那么要找到这个稀有变体需做许多筛选，使作图工作非常繁琐。而有些被称为可变数串联重复序列的 RFLP 显得更有用。它们源于小卫星系列，即一段 DNA 所含的核心序列一前一后地（首尾相连）重复很多次。不同个体某一 VNTR 内核心序列重复的次数可能不同，因此 VNTR 具有很高的多态性，作图也比较容易。然而，VNTR 作为遗传标记也有缺点，它们倾向于在染色体末端串联，而在染色体内部区域相对缺乏这种标记。

序列标签位点 另一种对基因组作图非常有用的标记技术为**序列标签位点**（sequence-tagged site，**STS**）。STS 是一类 60～1000bp 的短序列，可用 PCR 进行检测。图 24.9 说明怎样用 PCR 检测 STS。首先应非常清楚作图区的 DNA 序列并设计引物，上下游引物结合位点相距几百个碱基对，这样在两个引物间可以扩增出预计大小的 DNA 片段。用这对引物 PCR 扩增未知 DNA，如果出现预计大小扩增片段，则未知的 DNA 含有该 STS 位点。须注意的是仅引物与未知 DNA 结合还不够，引物结合位点一定要相距特定数目的碱基对并扩增出正确大小的片段。这就提供了一个特异性的结合位点可供参照。STS 作为制图工具的一个显著优势是不需要克隆、检测或冷藏 DNA。任何人只需根据发表的 STS 的引物序列合成同一引物并进行 PCR，就可以在短短的几小时里找出相同的 STS。STS 的另一优势是相对 Southern 印迹而言，PCR 反应所需 DNA 量要少得多。

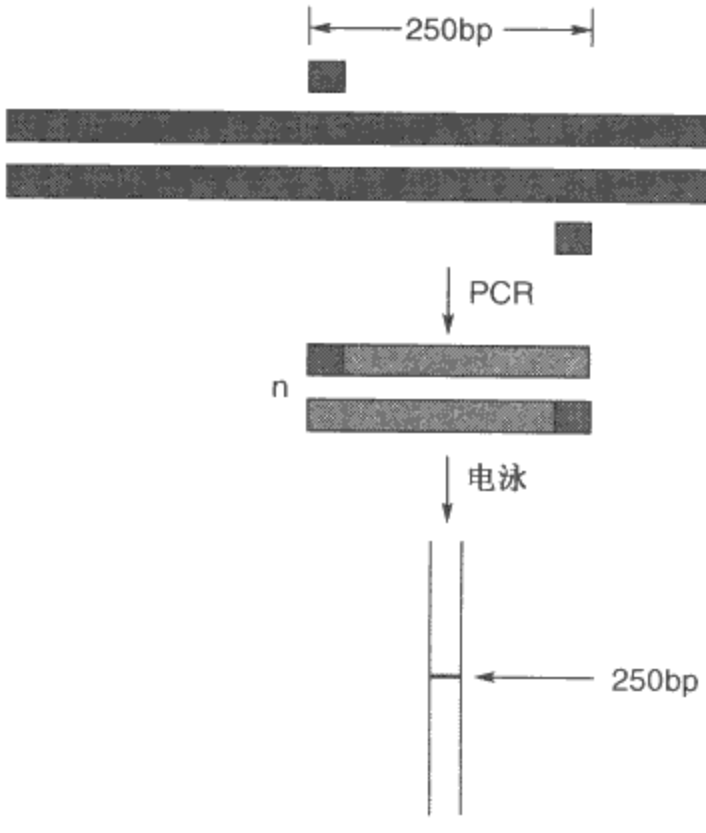


图 24.9 序列标签位点。以一个两端无限延长的 DNA 大片段为基础，我们知道其中一小段的序列，就可以设计引物，用 PCR 方法扩增出预计的片段。此例中，两引物（红色）间的序列长度为 250bp，多个循环后产生许多 250bp 的拷贝，可用电泳方法进行鉴定。

微卫星序列（microsatellite） STS 在物理作图和定位特定基因方面非常有用，但对传统遗传作图中的标记却毫无用武之地，除非具有多态性。只有具有多态性的 STS 才能用于遗传连锁分析。幸运的是遗传学家发现了一种被称为微卫星序列的 STS，它具有很高的多态性。微卫星序列（microsatellite）与小卫星序列（minisatellite）的相似之处在于它们均含有连续的核心重复序列。但典型小卫星的核心序列为 12bp 或更长些，而微卫星的核心序列比较小，通常为 2～4bp。1992 年，Jean Weissenbach 及

同事基于含有 C-A 二核苷酸重复的 814 个微卫星序列绘制了整个人类基因组的连锁图谱。他们分离了含有这些微卫星序列的 DNA，并根据 DNA 序列设计每个重复区两侧的引物。特定引物产生的 PCR 产物的大小取决于特定个体在该位点的 C-A 重复次数。幸好个体间微卫星序列的重复数差异很大，在人类基因组中，微卫星除具有多态性外，还具有广泛存在及分布均一的特点，因此微卫星标记法是连锁与物理作图的理想选择。

利用微卫星绘制遗传或连锁图与第 1 章提到的传统的果蝇遗传标记其实是同一技术。只是遗传学家是确定两个微卫星序列间的重组率而不是翅形和眼色间的重组率。例如，假设一男士在一个位点产生 78bp 的微卫星，并在附近一个位点产生 42bp 的微卫星，他的妻子在上述两个位点产生的微卫星分别是 102bp 和 36bp。在一定范围内，子女配子中这两个标记越是表现为非亲本（例如，孙子在上述两个位点产生的微卫星序列分别是 78bp 和 36bp），两标记间的重组率就越高，且这两个标记在染色体上的距离也就越远。

对基因组给定区域物理作图和测序感兴趣的遗传学家的目标是收集叫做**重叠群**（contig）的一组克隆，重叠群包含跨越很长距离的相邻 DNA（实际上是重叠 DNA）。这很像智力拼图玩具，模块越大，拼起来越简单，因此拥有高容量 BAC 和 YAC 载体很有必要。假设已获得人基因组的 BAC 库，我们还需要一些方法来鉴定出作图区包含的克隆。有几种不同方法可以达到此目的，其一是用靶区特异标记探针与 BAC 库杂交筛选，但这一方法的可靠性受非特异性杂交的影响。更可靠的方法是在 BAC 库中搜寻 STS 标记。最好用两个相距几百或几千个碱基对的 STS 筛选的 BAC 库，这样较大跨度的 BAC 才能被筛到。

获得一定数量的阳性 BAC 后，通过另外一些 STS 筛选就可以开始作图了。如图 24.10 那样以重叠的方式将筛出的 BAC 排列起来，该组重叠 BAC 就是新的重叠群，这时可以开始对重叠群进行精细的作图甚至测序了。

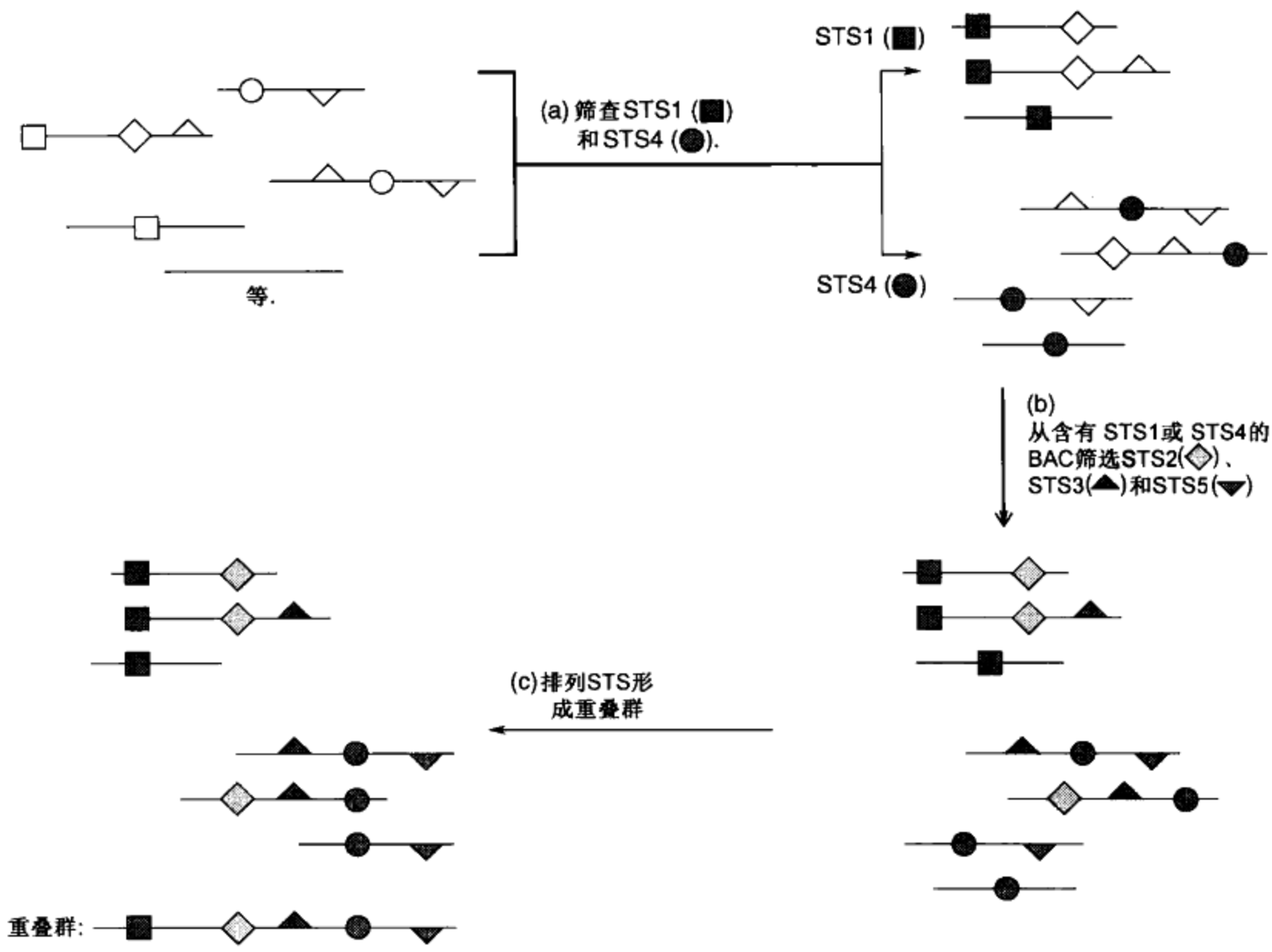


图 24.10　利用 STS 制图。左上侧为几个带有特定间隔的不同 STS（用不同的符号代表）的 BAC。第一步：筛选两个或多个相距较远的 STS，本例为 STS1 和 STS4，带有 STS1 和 STS4 的 BAC 在右上侧标出，已确定的 STS 标出不同颜色。第二步：经过第一步筛选出的 BAC，继续用 STS2、STS3 和 STS5 筛选。BAC 所标颜色为检测到的 STS。第三步：把 BAC 上的 STS 排列起来形成重叠群，用脉冲场凝胶电泳测量 BAC 的长度以帮助确定各 BAC 之间的相互联系。

辐射杂种作图法（radiation hybrid mapping）用 BAC 库作图听起来似乎很简单，其实还有很多困难。最困难是相对于人的整条染色体来说，BAC 的容量太小。构建整条染色体的 BAC 重叠群耗时费力，难以实施。所以我们需要一种新方法来发现间距超过单个 BAC 的 STS 间的连锁。辐射杂种作图就是这样一种方法。首先用致死剂量的电离辐射（如 X 射线和 γ 射线）照射人细胞，将人染色体打成碎片；将照射后的人细胞与仓鼠细胞融合，产生只含部分人染色体片段的杂种细胞；随后获得同一杂种细胞的克隆，即培育筛选一些细胞群，每一细胞群来源于单个祖先细胞；最后检测杂种细胞克隆中哪些 STS 倾向于在一起。越是常常在一起的 STS，相距越近，也越有可能位于同一染色体上。

1996 年，包括 G. D Schuler 在内的国际遗传学家联盟发表了基于该技术获得的 STS 绘制的人类基因组图谱。该图谱含有16 000多个 STS 标记，加上由传统的连锁分析法（家系研究）获得约 1000 多个遗传标记，构建了他们所公布基因组图谱的整体框架。这项研究中所用到的 STS 是一类被称为**表达序列标签**（expressed sequence tag，**EST**）的特殊标记，这些 STS 是从 mRNA 在反转录酶作用下合成的 cDNA 中获得的。上述 cDNA 可由 PCR 扩增和克隆获得，最后 cDNA 两端均被测序，产生两个小于 500bp 的序列标签。因此 EST 代表分离 mRNA 的细胞中所表达的基因。由于这种 STS（或 EST）分析获得的序列仅是一个基因的很小一部分，在 EST 数据库中一个给定的基因可能有许多不同的 EST。为减少 EST 的数量，作图联盟（the mapping consortium）规定仅能用代表基因 3′非翻译区的 EST 作图。这种方法的优势在于避免了大部分内含子区，而 3′非翻译区通常不含内含子区。1998 年，国际联盟（P. Deloukas 等）进一步细化和扩大了人类基因组图谱，其包括的基因数量已超过30 000个。

小结 人类基因组作图需要一系列路标，并建立基因位置与这些路标之间的联系。其中一些标记是基因，但更多的标记仅仅是 DNA 片段，如 RFLP、VNTR、STS（包括 EST 与微卫星序列）。其中 EST 和微卫星序列是能够以 PCR 扩增特定大小的 DNA 片段来鉴定的 DNA 区域。

鸟枪法测序

鸟枪法测序（shotgun sequencing）策略（图 24.11）由 Craig Venter、Hamilton Smith 和 Leroy Hood 于 1996 年首次提出。该方法绕过绘制基因组图谱，直接进行测序。首先对获得的已插入较大 DNA 片段（平均为 150kb）的一组 BAC 克隆进行测序，利用自动测序仪从每个 BAC 插入片段的两端进行测序，一次测序反应通常可以获得约 500bp 的序列，因此克隆（即插入片段）两端的 500bp 的序列可以测得。假设用这种方法对人类 DNA 的300 000个克隆进行测序，将测得 3×10^8 个碱基的序列，约为人类全基因组的 10%，在基因组中平均每 5kb 出现一个 500bp 的已测序区。对每个 BAC 克隆而言，这 500bp 序列作为一种识别标签（identity tag），称为**序列标签接头**（sequence-tagged connector，**STC**）。如果每个 BAC 克隆片段长度平均为 150kb，每 5kb 中有一个 STC，那么在 30 个 BAC 克隆（150kb/5kb＝30）中就有可能共享一个已知 STC，这是“接头”最初的含义，即每个克隆将通过其 STC 与其他 30 个克隆相连接。

紧接着是用限制酶消化每个克隆，做出其**指纹图谱**（fingerprint），主要有两个目的，一是鉴定插入片段的大小（限制酶消化产生的所有片段大小的总和）；二是排除异常克隆，即这些克隆的片段模式与通常意义上的重叠克隆不相符合。注意：该克隆的指纹图谱与基因组图谱不同，仅是测序之前的一个简单检测。

下一步是获得目的 BAC——**种子 BAC**（seed BAC）克隆的全序列。此过程是将一个 BAC 克隆细分为一组小片段的克隆，常常是在 pUC 载体中插入平均仅为 2kb 左右的片段。利用此 BAC 克隆的全序列能够鉴定出约 30 个与种子 BAC 重叠的 BAC 克隆，它们具有种子 BAC 所含的 STC。

接下来，在以上 30 个克隆中选择与第一个种子 BAC 重叠最少的进行测序，然后针对第二套 BAC 克隆重复上述过程选择与第二个测序的 BAC 重叠最少的克隆进行测序，并不断重复以上程序。这一策略被称作**BAC 步移**（BAC walking），如果有充足的时间，原则上一个实验室可以利用此项技术对人类全基因组进行测序。

但谁也没有那么多时间，于是 Venter 及同事改进了这一程序，他们随机测定 BAC 序列直至获得大约 35 亿个碱基。原则上这一序列比人类基因组大 10 倍以上，应该具有高度的覆盖率和准确率。接着将这些碱基序列输入计算机中，其强有力的程序可发现两个克隆间的重叠并将这些序列拼接起来，形成全基因组序列。

如前所述，大批量测序用的是 pUC 载体，其插入片段相对较小（仅约 2kb）。这些小插入片段不能提供足够的重叠来拼凑全基因组。这一缺点在具有重复序列的 DNA 区尤为明显。从一个具有**串联重复**（tandem repeat）的 10kb DNA 区域获得一个 2kb

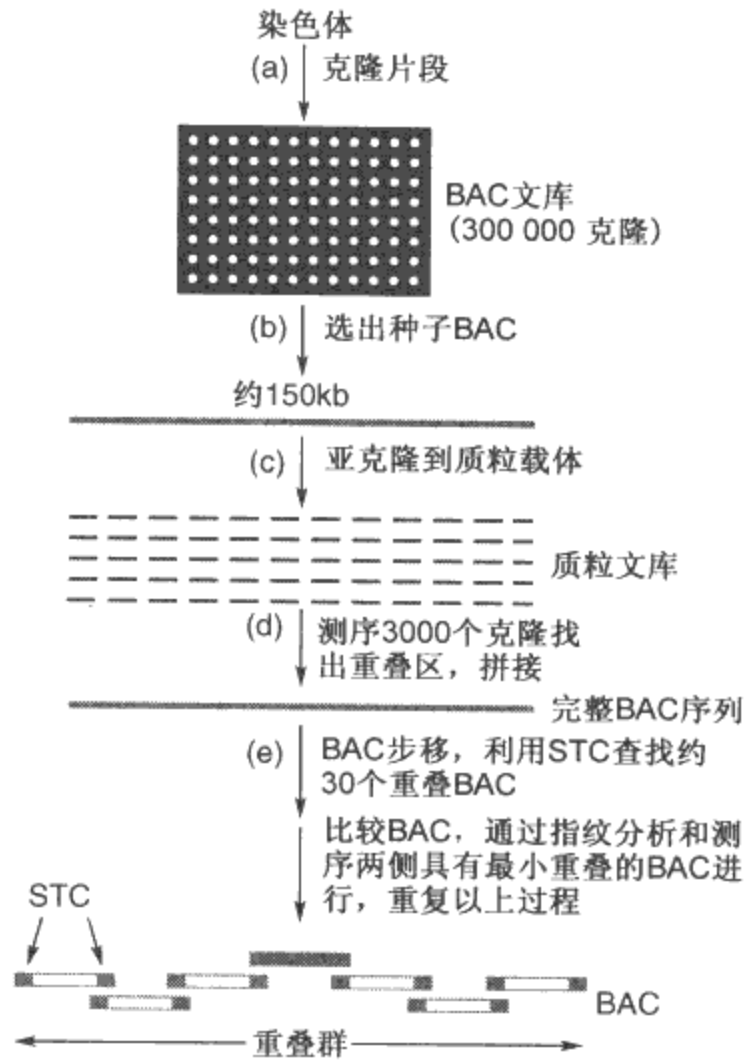


图 24.11　鸟枪法测定序列的方法。（a）染色体被克隆到 BAC 载体上，产生 300 000 个 BAC 克隆。每 96 孔板中有 96 个克隆。（b）选择一个种子 BAC 克隆（seed BAC）进行序列的测定。（c）种子 BAC 被亚克隆到质粒载体上。（d）3000 个质粒克隆被测序，通过序列重叠进行排序，最终获得 150kb 的 BAC 克隆全序列。（e）寻找具有重叠 STC 的 BAC（约 30 个），通过指纹图谱对它们进行比较，找出其中具有最小重叠的克隆并测序。这一过程称作 BAC 步移，原则上创建了一个覆盖全基因组的叠加群。（*Source*：Adapted from Venter，J. C.，H. O. Smith and L. Hood，A new strategy for genome sequencing *Nature* 381：365，1996.）

序列，并不能确定它到底位于大重复片段的哪个位点，因为大片段的局部序列可能是相同的。而这正是 BAC 克隆能派上用场的地方，它们足以覆盖几乎任何重复区域，也提供跨越较大 DNA 区域的重叠序列，因此，有助于将较小的克隆片段组织到一起。物理图谱，尤其是 STS 图谱也促进了这项工作，这些图谱早已获得。所以，鸟枪法测定人类基因组序列在实践中是由纯粹鸟枪法与图谱-测序结合的一种策略。

对超过 30 亿 bp 进行测序的任何策略都必须依赖于高容量、低成本的测序方法，我们现在所用的测序装置是在毛细管中进行 DNA 片段电泳，而不是传统的薄凝胶板电泳。这些仪器是完全自动化的，每台每天能处理 1000 个样品，只需 15min 的人工耗时。Venter 小组的另一个公司基因组研究所（The Institute for Genome Research，TIGR）有 230 台这样的仪器，每天他们能够以较低的劳动成本同时测出 100Mb 的 DNA 序列。

小结　大范围的测序计划可采取两种形式：①先作图再测序策略（map-then-sequence strategy），先构建一个包括 STS 的基因组物理图谱，然后对作图过程中用到的克隆（主要为 BAC）进行测序，这样就可以按正确顺序将这些序列拼凑在一起；②在鸟枪法中，先构建一个含有不同大小插入片段的克隆文库，然后随机测定这些插入片段的序列，这种方法必须依赖于计算机程序去发现序列之间的重叠区，然后把它们拼接在一起。在实践中，这些方法被组合在一起用来测定人类基因组序列。

测序的标准

基因组“草图”（rough draft）、“工作草图”（work draft）和“完成草图”（final draft）的含义是什么？这要看你问谁了。多数研究者认为工作草图仅能完成 90%，而且可能还有约 1%的出错率。尽管对完成草图还没有一个统一的量化标准，但大家一致认为它的出错率应小于 0.01%，而且间隙应尽可能少。一些分子生物学家坚持认为，人类基因组不可能被完全测序，除非每个间隙的序列都填满，但这很困难。我们将在下一节提到，由于许多未知的原因，DNA 中的一些区域很难被克隆。克服困难去克隆那些区域的花费十分高昂，所以人类基因组的最后几百万碱基的填充任务是无法完成的。在下一节将详述人类基因组 22 号染色体测序组织认为他们所测定的序列只是“功能性地完成”，利用目前所有的克隆和测序工具他们已经获得了所有可能获得序列，但仍然存在着明显的间隙。

人类基因组测序

1999 年底，我们品尝到了人类基因组计划的第一个果实：人类 22 号染色体的完成草图。2000 年

中期，国际联盟公布了人类 21 号染色体的完成草图，虽然这是人类最小的两个常染色体，但它们的序列很有价值。2001 年 2 月，Venter 小组和国际联盟组织各自公布了他们的人类基因组工作草图。2004 年，国际联盟宣布完成了人类基因组常染色体的部分测序。

本节将介绍完成测序的两个最小染色体、X 染色体及全基因组工作草图所获得的信息。在开始之前需要注意，所有 3 个染色体的序列都是基于更有序的逐步克隆法。采用这一策略，一旦各小组完成对染色体的测序，就可获得完整染色体的完成草图序列。而利用鸟枪法测得的原始序列必须由计算机发现重叠序列后才能建立重叠群进行拼接。因此该方法在整个基因组测序完成前，得不到染色体的最终草图序列。

22 号染色体　实际上只测定了染色体长臂（22q）的序列，短臂（22p）由纯异染色质组成，被认为是基因空白区。而且，在序列中还存在 11 个间隙，其中有 10 个是重叠群间的间隙，这些间隙无法填充——推测是无法克隆的 DNA。还有约 1.5kb 的已克隆 DNA 区域测不出序列，一些 DNA［有时被称为“有毒区域”（poison region）］不能被克隆的原因还不完全清楚。但有一点很清楚，具有异常二级结构和重复序列的 DNA 常从细菌细胞中丢失，这就是异染色质（第 13 章）区难以测序的原因之一。异染色质主要存在于染色体的着丝点和端粒附近，并且富含重复序列。虽然未能获得基因组异染色质的序列，科学家们并没有丢失太多基因，因为一般认为此处不存在基因。但是可能会错过这些异染色质区的一些令人感兴趣的东西。

22 号染色体序列提供了什么信息呢？有几个令人兴奋的发现。首先，我们不得不接受人类基因组序列中存在间隙的事实，虽然可能并不像第一次测序时在该染色体中出现的那么多。截至 2000 年夏季，一个间隙已被填充，更多的间隙可能会及时被填充。其他研究人员在测定其他染色体序列的过程中，也遇到了类似的间隙问题。表 24.2 列出了 22 号染色体中已测序列的重叠群和其间的间隙。重叠群33 464kb，是染色体的 97%，且已测出的序列有较高的精确率，估计每 50 000 个碱基中不到一个错误。有趣的是所有间隙都出现在染色体的着丝粒和端粒附近。间隙 4 和 5 之间是一个巨大的重叠群，由23 006kb个碱基组成，覆盖了 22q 染色体的 2/3 还多。

第二个主要发现是 22 号染色体上包含 679 个**注释基因**（annotated gene）（至少部分被鉴定的基因或类基因序列），可分为以下几类：**已知基因**（known gene），其序列与已知的人类基因相同或与已知的人类蛋白质推测出来的序列相同；**相关基因**（related gene），其序列与已知的人类基因或其他物种同源，或与已知基因有相似的区域；**预测基因**（predicted gene），含有与 EST 序列同源的序列（所以确信它们是被表达的基因）；**假基因**（pseudo-gene），其序列与已知基因同源，但含有阻碍适当表达的缺陷。在染色体 22q 中有 247 个已知基因、150 个相关基因、148 个预测基因和 134 个假基因。因此，不计假基因，仍有 545 个注释基因。计算机序列分析还预测了另外 325 个基因，但这些预测很不精确，因为运算法则依赖于发现外显子，而人类基因组的许多长内含子使得外显子很难被辨认。在计算机预测的 325 个基因中，最好的预计是有约 100 个真实基因，所以可以估计人类 22 号染色体约有 645 个基因。

第三个主要发现是，基因的编码区仅占染色体长度的很微小部分，即使算上内含子，注释基因仅占 22q 全长的 39%，外显子仅占 3%。通过对比，22q 中的 41% 是重复序列，尤其是 Alu 序列和 LINES（第 23 章）。表 24.3 列出了 22 号染色体上的分散的重复序列及其重复率。

表 24.2　22 号染色体上的重叠群和间隙

重叠群	间隙	片段大小/kb
1		234
	1	1.9
2		406
	2	≈150
3		1 394
	3	≈150
4		1 790
	4	≈100
5		23 006
	5	≈50
6		767
	6	≈50～100
7		1 528
	7	≈150
8		2 485
	8	≈50
9		190
	9	≈100
10		993
	10	≈100
11		291
	11	≈100
12		380
总的序列长度		33 464
22q 的总长度		34 491

（*Source*：Adapted from Dunham, I., N. Shimizu, B. A. Roe, S. Chissoe, A. R. Hunt, J. E. Collins et al., The DNA sequence of human chromosome 22. *Nature* 402：491，1999.）

第四个主要发现是染色体上不同区域重组率变化很大。重组率相对较低的长区域散布于重组率相对较高的短区域（图 24.12）。在该染色体上，遗传距离与物理距离的比率是 1.87cM/Mb。当然，应记住 y 轴表示累加的遗传距离，是相距很近的标记之间距离的总和。相距较远的标记间的实际距离并不是其内部标记间距的总和，因为多次重组事件较有可能发生在距离较远的标记之间，因而使这些较远的标记显得比其真实距离要近一些（第 1 章）。

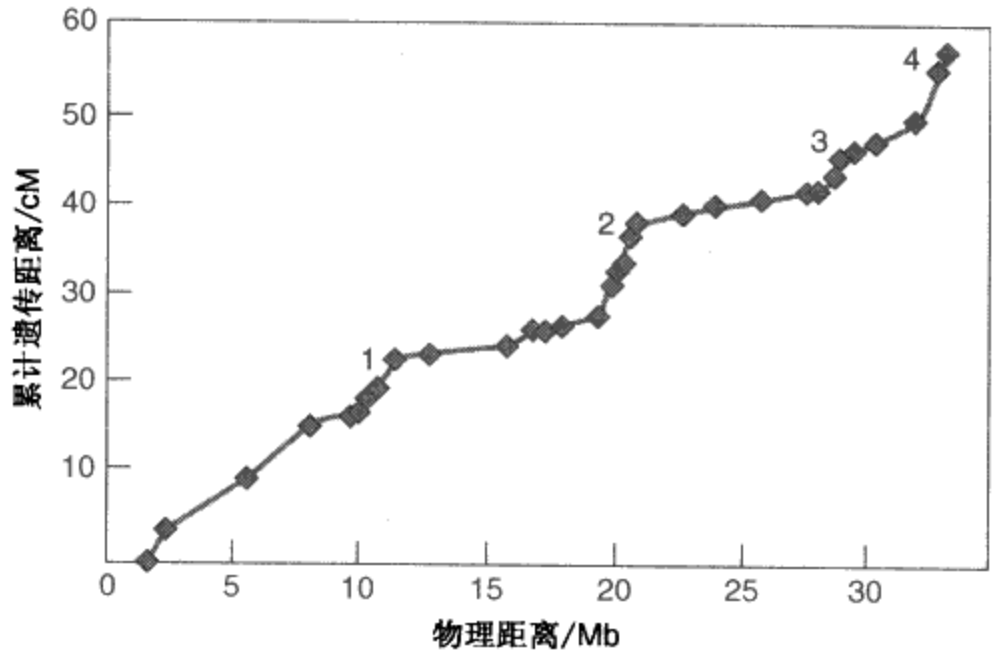


图 24.12　22 号染色体长臂上的遗传距离对物理距离的作图。标记之间累计遗传距离（cM）对物理距离（Mb）作图。数字指明了 4 个重组率相对较高的区（由曲线的陡峭部分所反映）。（*Source*：Adapted from Dunham, I., N. Shimizu, B. A. Roe, S. Chissoe, A. R. Hunt, J. E. Collins et al. (The Chromosome 22 Sequencing Consortium), The DNA sequence of human chromosome 22. *Nature* 402: 492, 1999.）

第五个主要发现是 22q 染色体上有几个局部的、长距离重复。最明显的是免疫球蛋白 λ 的基因座。这个基因座上有 36 个基因片段成簇聚集在一起，这些基因片段至少具有潜在编码 λ 可变区（V-λ 基因片段）的能力，还有 57 个 V-λ 假基因和被称作“遗迹”的 27 个部分 V-λ 假基因。其他的重复被很长的片段分开。一个极端的例子是，60kb 区域在 12Mb 以外出现保真度大于 90%的重复区。与分散的重复序列（如 Alu 序列和 LINES）比较，此类重复的拷贝数较低，被称为**低拷贝重复**（low-copy repeat, LCR）。前面所描述的 22q 着丝粒末端的 8 个 LCR22 中的 7 个已被测序，第 8 个（LCR22-1）可能位于最接近于着丝粒的间隙中。

第六个主要的发现是人类 22q 染色体的几个大区段在小鼠的几个不同染色体上存在同源区域。测序组发现 113 个人类基因的小鼠**直系同源基因**（ortholog）已经在小鼠染色体上定位［直系同源基因是不同物种间的同源基因，来自于共同的祖先基因；**旁系同源基因**（paralog）是一个物种内由基因复制进化产生的同源基因；**同源基因**（homolog）是任何类型的同源基因，包括直系和旁系同源基因］。小鼠的直系同源基因分别成簇聚集在 7 个不同染色体的 8 个区域上，如图 24.13 所示。呈现人 22 号染色体长臂上同源区段的小鼠染色体有：5、6、8、10、11、15 和 16。小鼠 10 号染色体在两个区域呈现人类 22q 的同源基因。在两个物种进化过程中发生分叉，染色体已经被重排了，但是许多标记间的连锁被保留在**同线群**（syntenic block）中。很明显，对于人类基因组序列的了解极大地促进了小鼠基因组

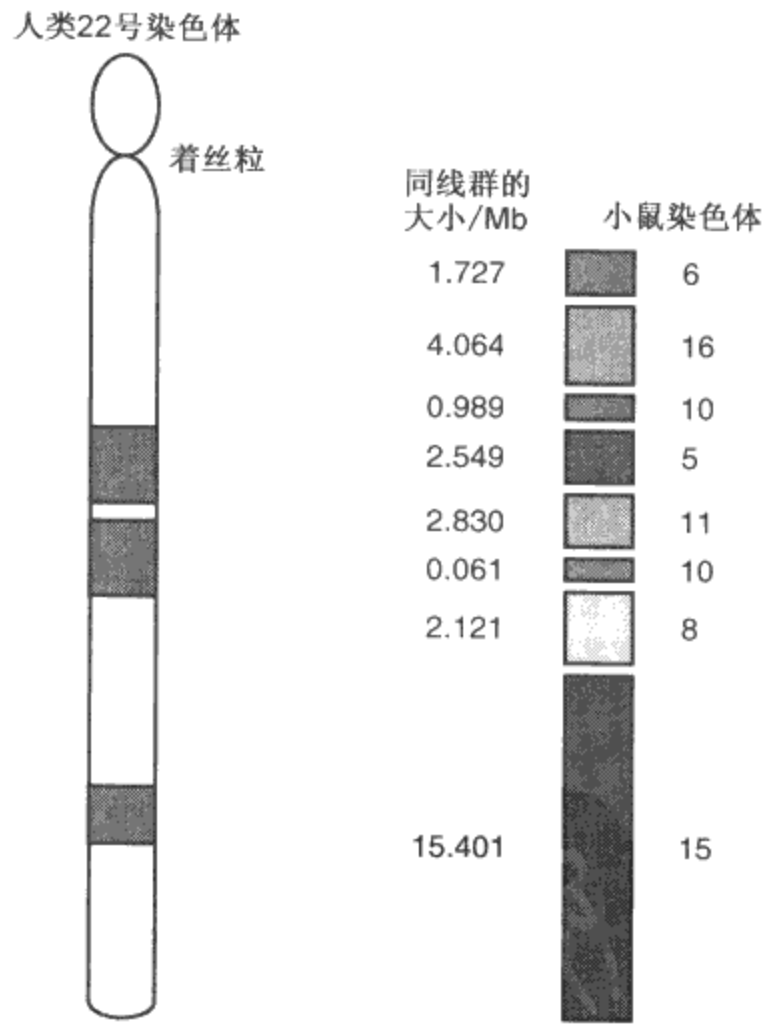


图 24.13　人和小鼠染色体上的保守区。人类 22 号染色体如左图所示，着丝粒靠近顶端，主要区带以白色和棕色表示。含有同线群的 7 个不同小鼠染色体如右侧所示。与小鼠染色体对应的颜色在最右边列出。（*Source*：Adapted from Dunham, I., N. Shimizu, B. A. Roe, S. Chissoe, A. R. Hunt, J. E. Collins et al. (The Chromosome 22 Consortium), The DNA sequence of human chromosome 22. *Nature* 402: 494, 1999.）

的测序。

小结 人类 22 号染色体的长臂已被精确测序，但仍有 10 个间隙不能用目前的方法填充。有 679 个注释基因，但染色体主要由非编码 DNA 组成，40%以上的非编码 DNA 散布在重复序列、Alu 序列和 LINES 中。染色体上不同区域重组率变化很大，低重组率的长区域被高重组率的短区域所打断。染色体包含一些局部的长距离间隔重复序列。人的染色体包含有某些大的区域，其基因间连锁区段在小鼠的 7 个不同染色体中都能找到相应的同源区。

21 号染色体 尽管染色体编号倒数第二，21 号染色体实际上比 22 号还要小一点，但遗传学家对这条染色体特别感兴趣，因为它与唐氏综合征（Down syndrome）有关。正常人有 2 个 21 号染色体，而唐氏综合征患者有 3 个 21 号染色体（trisomy 21），该病造成了患者智力和身体的异常。从 21 号染色体序列上能了解到什么？与 22 号染色体比较有何区别？

同样，长臂（21q）是 21 号染色体测序的主要靶区，因为短臂仅含很少的基因，且长度在不同人之间有很大变化。但测序联盟还是对短臂的 281kb 进行了测序，并且在大量高度重复序列中发现了至少一个基因。与 22 号染色体类似，21 号染色体的序列中也存在间隙，但只有 3 个，全部合起来仅 100kb。这意味着 21 号染色体被测定的全部序列（33.55Mb）比 22 号的 33.46Mb 稍多一些。

通过序列差异所揭示的两个染色体之间的最明显的不同是 21 号染色体的基因密度相对要低些。测序联盟在 22 号染色体发现了 545 个基因，而 21 号染色体只发现了 225 个基因，这些基因中的 127 个是已知基因，98 个是预测基因，还鉴定出 59 个假基因。这种较大差异可以部分地解释为两个测序组用于定位基因的技术不同，22 号染色体测序组部分地依据与 EST 的匹配性鉴定基因，而 21 号染色体测序组依据更严格的标准，即与染色体一定区域匹配的 EST 必须显示从一个较大前体拼接过的标志，就像大多数人类转录产物一样。所以，这种比较保守的方法可能低估了 21 号染色体的基因数。

表 24.3 人类 22 号染色体重复 DNA 的含量

类型	总数/个	总的碱基对数/bp	在染色体所占比率/%
Alu	20 188	5 621 998	16.80
HERV	255	160 697	0.48
LINE 1	8 043	3 256 913	9.73
LINE2	6 381	1 273 571	3.81
LTR	848	256 412	0.77
MER	3 757	763 390	2.28
MIR	8 426	1 063 419	3.18
MLT	2 483	605 813	1.81
THE	304	93 159	0.28
其他类型	2 313	625 562	1.87
二核苷酸	1 775	133 765	0.40
三核苷酸	166	18 410	0.06
四核苷酸	404	47 691	0.14
五核苷酸	16	1 612	0.0048
其他串联形式	305	102 245	0.31
总和	55 664	14 024 657	42.91

Source: Adapted from Dunham, I., N. Shimizu, B. A. Roe, S. Chissoe, A. R. Hunt, J. E. Collins et al., The DNA sequence of human chromosome 22. *Nature* 402: 491, 1999.

此外，21 号染色体联盟测序组采用了更灵活的计算机运算法则［计算机模拟预测（in silico prediction）］评估基因，而 22 号染色体测序组在校正时必须从计算机模拟预测中去掉假阳性，21 号染色体组没有进行校正，有可能高估了 21 号染色体的基因数。所以，21 号染色体测序小组使用的两种方法，一种保守，另一种灵活，可能相互抵消一部分。因此，可以比较有把握地得出结论：21q 染色体相对于 22q 染色体的每兆碱基所含基因较少。这也许有助于解释为什么细胞允许一个额外的 21 号染色体拷贝，尽管它导致了唐氏综合征。基因数目较少可能降低了有一条额外染色体拷贝的个体的致死率。

可以用两个染色体中估计的基因数预测人类基因的总数吗？这种推测是有风险的，因为是以整个人类基因组的2%为基础。21号染色体联盟测序组的预测是：人类基因组的基因少于40 000个，这一数字少于遗传学家曾经预测的100 000个，但仍然高于根据已完成的人类全基因组草图所估计的23 000个基因。

在21号染色体序列中发现的另一个有趣现象是小鼠10号染色体的24个基因与人21号染色体的基因同源，且这些基因以相同的顺序出现在两个染色体中。这种同源性是两个物种进化时遗传信息高度保守的另一个标志。

21号染色体的序列测定无疑促进了遗传病相关基因的研究，不仅仅是唐氏综合征，还包括21号染色体的单个基因缺失所造成的疾病，如阿尔茨海默病（Alzheimer's disease）和肌萎缩性脊髓侧索硬化症（Lou Gehrig's disease）。

小结 人类染色体21q和小部分21p已经测序，剩下的间隙相当少并且不长。序列揭示该染色体基因较少，仅约225个，另有59个假基因。小鼠10号染色体和人21号染色体共有的24个已知基因以相同的顺序出现在两个染色体中。

X染色体 X染色体之所以尤为引人注意是其在性连锁疾病中的特殊地位，这些疾病包括血友病（hemophilia）和杜兴氏肌营养不良症（Duchenne muscular dystrophy）等。男性只有一条X染色体，在X染色体上的基因若发生突变，由于缺少相应的野生型等位基因作“弥补”，所以很容易表现出病态。另外，X染色体的进化也同样极具吸引力，X和Y染色体源于一对常染色体（非性染色体），经过长时间的进化，X和Y染色体的同源区几乎无法识别了。在分化过程中伴随着Y染色体基因的丢失，两个性染色体只能在一个小的同源区内结合在一起。

2005年，Mark Ross和许多合作者公布了人类X染色体的完成序列。对其155Mb的全长测序了151Mb，超过该染色体常染色质的99.3%。他们发现该染色体的基因密度很低，只有1798个注释基因，其中包括700个假基因。而22号染色体在34Mb的DNA序列中含679个注释基因。另外，X染色体富含L1元件（L1 element）。该元件的确发挥着重要的作用，我们随后将看到相关内容。

对X染色体上的基因是如何进行注释的呢？与描述人类21号和22号染色体的作者相比，描述X染色体的作者的确有一个很大优势，因为另外5种脊椎动物［小鼠（mouse）、大鼠（rat）、斑马鱼（zebrafish）、两种河豚（pufferfish）］的X染色体测序已完成。通过对6个物种X染色体序列的比对，从人类X染色体上发现了4493个**进化保守区**（evolutionarily conserved region，**ECR**）。这些ECR主要位于基因的外显子中，尽管有少数几个可能是高度保守的调控区和结构区。4493个ECR中有4393个被注释为位于编码蛋白质的外显子区，可以保守地假设这4393个ECR全部位于外显子区，这表明作者已注释了人X染色体97.8%（4393/4493）的外显子。如果对人类所有染色体进行上述分析，可更精确地估计人类基因组所含基因的总数，现在认为这个数目在20 000～25 000。

X染色体还编码173个非编码RNA（ncRNA），其中一些RNA的功能很难注释，因为基因和假基因间较难区别。对于这些非编码RNA，只有当其转录物在cDNA文库或表达序列标签中得以确认才能编入基因名录中。该染色体仅包含两个tRNA基因，与预测的全基因组中总共有几百个tRNA基因相差甚远。X染色体还编码了13个microRNA（第16章），可在miRNA注册网站找到这些分子。

最引人注目的ncRNA可能是32kb的XIST RNA｛X（非活性）特异性转录物RNA［X（inactive）-specific transcript RNA］｝，它在女性X染色体失活（X chromosome inactive，XCI）中起重要作用。在胚胎早期，在女性个体中为了使其X染色体的基因剂量与只有单个X染色体的男性保持平衡，一条X染色体上的大多数基因保持沉默，这种沉默在这些细胞及其分裂后代细胞中持续发生。XIST RNA通过覆盖X染色体而导致基因沉默，这一事件一旦发生，X染色体上仅有少数几个基因能够逃避沉默，*XIST*基因就是其中的一个，它必须保持活性以使这种沉默无限期地延续下去。但沉默只是顺式（*cis*）的，即只在同一条染色体上发挥作用。因此，*XIST*基因在有活性的X染色体上是关闭的。

为使整个X染色体保持沉默，需要许多“驿站”（way station）来传递沉默信息，有关驿站的早期证据源自当其他染色体的片段转座到雌鼠X染色体后，转座的染色体片段转为沉默，但只是一定程度的沉默，可能是由于这些片段中缺乏驿站。1998年，Mary Lyon提出L1元件可能起驿站作用，X染色体的序列支持这一假说。相对而言，X染色体富含L1元件，L1元件遍及整个染色体。对X染色体不同区域分析发现，L1元件数与基因沉默的程度密切相关。

X染色体的序列还揭示了X染色体的演化。前面提到X和Y染色体源于一对常染色体，通过对人X染色体序列与鸡基因组序列的比较，有力地支持了这一假说。人类X染色体长臂的大部分区域与鸡

4 号染色体（为常染色体）短臂末端 20Mb 的片段同源，而人类 X 染色体短臂的大部分区域与鸡另一个常染色体（1 号染色体）的片段匹配。因此，在人和鸡共同祖先的一个或多个染色体中都含有这一批基因。对鸡而言，这个或这些染色体当然是进化了，只是它们仍然是常染色体，而在人类这些染色体重组到一对染色体上（如果它们之前是分开的），并且进化为一对性染色体。

X 染色体进化的另一方面表现为哺乳动物 X 染色体保持高度保守的同线性（synteny），小鼠 X 染色体有 9 个较大区域的序列与人 X 染色体同线（syntenic），而大鼠 X 染色体有 11 个这种区域。这些同线区几乎布满整个 X 染色体。为什么会保持这么高的同线性呢？Susumu Ohno 提出 X 染色体失活机制，该机制一旦建立，染色体不再发生显著的重排，所以不同生物的 X 染色体基因倾向于保持相同的顺序，有胎盘的哺乳动物正是如此。

最后，本节开始时提到 X 染色体所含的引起人类疾病的基因与其基因总数不成比例。虽然 X 染色体基因相对贫乏，仅占人类基因总量的 4%左右，但在所有单基因遗传病（mendelian disorder，又称为经典遗传病）中，10%的疾病相关基因定位于 X 染色体上。因此，X 染色体的序列必定有助于医学研究。迄今为止，图位克隆已经确定了 168 种性连锁的表型，其中 43 种得益于 X 染色体序列的测定。

小结 人类 X 染色体 151Mb 序列的测定（常染色质的 99.3%）揭示了 1798 个蛋白质编码基因（包括 168 个与性连锁表型相关的基因）和 173 个非编码 RNA 基因。X 染色体富含 LINE1 元件，该元件在女性 X 染色体失活机制中可能起驿站作用。人类 X 染色体序列和鸡基因组序列之间的比对结果证实了 X 染色体（以及 Y 染色体）源于一对常染色体。3 种哺乳动物 X 染色体序列的比对证明这些染色体保持着高度保守的同线性，可能反映了进化压力稳定地保持了这些基因在 X 染色体上的顺序。

人类基因组工作草图和完成版本 2001 年 2 月，Venter 小组和国际联盟分别公布了他们各自的人类全基因组工作草图的版本。当然，这两个版本并不完善，存在许多缺陷和错误，但是它们包含了大量有用的信息，科学家们要花很长时间对其分析和补充。此外，随着各独立测序小组完成其艰难的最后阶段，即填补缺口、更正错误，这一公布的草图还将继续完善。

这两个组最惊人的发现是基因组的基因数很低。Venter 小组发现 26 588 个基因（至少有两方面的证据支持）及 12 000 多个潜在基因。这些潜在的基因只是根据计算机的数据确定的，还没有其他的数据支持这一猜测，他们推测这些潜在的基因序列大部分为假阳性。而国际联盟预计人类基因组包含有 30 000～40 000 个基因。我们随后将看到，根据已测定的人类基因组序列估计还要再少一些，应该少于 23 000。

因此，与早期估计相反，人类基因的数目似乎只有低等蛔虫或果蝇基因的 2 倍。显然，一种生物体的复杂性与它所含基因的数量不成正比。那么怎么解释人类的复杂性呢？一个现成的解释是人类基因的表达比简单生物体要复杂。例如，估计至少有 40%的人类基因组转录物经历了选择性剪接（第 14 章）。因此，较少数量的编码蛋白质结构域和基序的基因序列可以不同方式组合产生具有不同功能的丰富多样的蛋白质。另外，人类蛋白质的翻译后修饰似乎比简单生物更为复杂，这也是产生人类蛋白质功能多样性的原因。

另一个重要发现是约有一半的人类基因组源于转座元件，这些元件在基因组中自主复制并携带人类 DNA 从一个位置转移到另外一个位置（第 23 章）。虽然转座子对基因组有巨大的贡献，但很大一部分转座子仍处于非活性状态。事实上，一方面所有的 DNA 转座子都是无活性的，所有包含 LTR 序列的反轨录转座子也无活性；另一方面，如第 23 章所介绍的，人类基因组有少数 L1 转座子仍有活性，并引发一些人类疾病。

许多人类基因可能源于细菌，其他的来自进入人类细胞的新转座子。因此，人类基因组的形成不全是由于体内基因的突变和重排，还包括外界基因的引入。

经过多年的预测，人类基因组的全长约30 亿 bp（3Gb）。Venter 小组测序结果为 2.9Gb，国际联盟预测基因组的全长约为 3.2Gb。

正如本章前面所提到的，测定人类基因组的国际研究协会在 2003 年春季宣布他们已经完成了人类基因组的测序工作，比原计划提前了两年。他们于 2004 年发表了工作结果，与草图相比这一版本的主要优点如下。

1. 该版本更完全。99%的可获得序列已经获得，测序结果为 2 851 330 913bp 或 2.85Gb（即 28.5 亿）。

2. 该版本更准确。错误率仅 0.001%，并且所有序列都被正确拼接。

然而，仍然有 341 个缺口，尽管其中的 33 个是在人类基因组计划靶区之外的异染色质区。当然，生物学家一般都承认我们得容忍存在缺口。再者，尽管对已完成的草案进行了修饰，但注释仍然很困

难，我们仍然不知道人类基因组中真正的基因数。国际联盟发现了 22 287 个编码蛋白质的基因，包括 19 438 个已知基因和 2188 个预测基因，比基于两张人类基因组草图做出的估计少得多。出现这一差异主要是由于早期（基因组作图阶段）对同一个基因多次出现的重复计数。

我们还能从完成的基因组图谱中了解什么呢？以下是几个例子，22 289 个基因出现 34 214 个转录物，大约每个基因出现 1.5 个转录物。这些基因表现为 231 667 个外显子，即每个基因有 10.4 个外显子，包括所有这些外显子的 DNA 数量仅仅是 34Mb，这仅仅是人类基因组常染色质的 1.2%。这些信息证实了我们已经知道的一些知识，即人类基因组的大部分不包含蛋白质编码基因，而是编码 rRNA、tRNA、snRNA 和 miRNA，当然，这些 RNA 不被翻译。但其中的一部分甚至根本不被转录，它们的功能（如果有的话）仍然是个谜。

完成的基因组图谱也有助于人类进化的研究，首先，它重新揭示了基因成对出现的机制，即在原始基因的基础上产生具有新功能的新基因，一对基因中的一个保持它的原始功能，而另一个可以不断地突变并演化出新的活性，这不会危及它的原始功能，对生存而言也是必需的。

其次，完成的基因组图谱揭示了新近失活的基因或假基因。假基因的研究始于大鼠、小鼠和人类基因组的比较，结果发现在 3 种生物中有成串的基因，接着研究者从这些成串基因中寻找只存在于啮齿动物而不存在于人的基因。最后，他们在人类基因组检测预计含有这些丢失基因的区域，结果发现了 37 个候选假基因，这些基因虽然都已失活但仍可清晰地识别。每个假基因平均有 0.8 个非成熟终止密码子、1.6 个移码（frameshift）。这些突变形式的任何一个都导致了基因的失活。显然，这些基因不是人类生存所必需的，虽然它们可能是人类、大鼠和小鼠（或者可能还是啮齿动物）共同的祖先。

为了证实这些假基因确实是失活的，研究者重新对其中的 34 个进行测序，其中 33 个确实是失活基因，另一个是测序出错造成的。随后他们将这 33 个序列与黑猩猩基因组中的相应序列进行比对，发现这些假基因中的 19 个有 2 个以上的失活突变，且在黑猩猩中也都是假基因，另外 14 个只有 1 个失活突变，有趣的是其中 8 个在黑猩猩中也是假基因，但另外有 5 个是有功能的基因，还有一个表现为多态型（即在一些个体中为假基因，而在另一些个体中为功能基因）。因此，我们可以通过进化的时代看到基因失活的痕迹——从啮齿动物与人类的分化以及黑猩猩与人类的分化。

小结 人类基因组草图已由两个独立的研究组报道，结果显示人类基因组所含基因可能比预计的更少，仅为 25 000～40 000 个。基因组的一半源于转座子作用，转座子为基因组贡献了许多基因。除此之外，细菌似乎也贡献了不少基因。已完成的人类基因组草图比工作草图更加完全和准确，但仍有一些缺口。基于完成的草图，遗传学家估计基因组包含 20 000～25 000 个基因。完成草图也在人类进化过程中基因的产生与消失方面提供了有价值的信息。

其他脊椎动物基因组

我们已经看到，其他脊椎动物（小鼠和河豚）的部分序列是如何帮助识别人类 20 号染色体基因的。现在，小鼠和东方红鳍鲀（tiger pufferfish）全部序列的草图已经公布了。从这些基因组中可获得什么信息呢？有以下几点比较重要。

对东方红鳍鲀基因组测序是因为其基因组比人类的小得多，仅为人类序列的 1/9。二者的不同之处不在于基因序列，而在于内含子的大小和 DNA 重复数的不同。东方红鳍鲀基因的内含子比人的要少得多，DNA 的重复序列也少得多。比较东方红鳍鲀和人基因组，基因组学家可鉴别出人的 1000 个基因。

导致人类疾病的基因突变很可能发生在基因的重要位点，并且由于这些重要位点非常保守，所以可通过比较亲缘关系较远的两组脊椎动物（如人和东方红鳍鲀）基因组来识别这些重要的位点。比较时，通常不用小鼠基因组，因为它和人类基因组比较相似，其原因只是小鼠和人类基因组尚无足够的时间分化得更远，除一些重要位点外，其他许多位点也很保守。

与 3Gb 的人类基因组相比，小鼠的基因组稍微小一点，约 2.5Gb，但二者有相同的基因数，而且绝大部分是相同的，小鼠 99%的基因可以在人类基因中找到其相应部分。剩下的 1%很难解释人类和小鼠之间明显的生物学差异，因此除了纯 DNA 之外，应该还有其他物质在起作用。初步研究表明，造成人与小鼠的不同可能不是基因本身而是基因的调控。认识到小鼠和人在基因组结构上很大的相似性，科学家用小鼠代替人，可做一些不能在人体进行的实验。例如，他们可以敲除小鼠中的基因，然后观察对其造成的影响，这提供了人类相应同源基因功能的线索。分子生物学家还研究了小鼠发育期及成体的基因表达方式，如这些基因在什么时候及在哪些部位表达。这些结果同样能给人类相应同源基因的表达方式提供有益的信息。

2003 年初，比较人与小鼠基因组的最好研究主

要是针对已经完成测序的染色体，包括人类 21 号染色体和小鼠 16 号染色体。下面分别介绍这些研究结果。

比较人类 21 号染色体和对应的小鼠 DNA，发现有 3000 个保守序列。更为惊讶的是仅有一半的保守序列含有基因。而这种高度保守性表明保守序列很重要。为什么会如此保守？也许它们在基因表达中起作用。人类有 234 个基因匮乏的所谓“基因沙漠”（gene desert），而其中 178 个在小鼠中非常保守，这种看起来无用的 DNA，其保守程度需要解释。接着，基因组学家们敲除了小鼠的一些基因，以研究这些基因的丢失会造成什么影响。

2002 年，Venter 及同事报道了人类基因组序列和小鼠 16 号染色体序列的比较。他们发现很多同线区域，即这些区域有很高的同源性，可能起源于相同的哺乳动物染色体。图 24.14 从蛋白质水平分析了这些同线区，发现小鼠 16 号染色体与人类 6 条染色体（以不同颜色标出）有同源性；在两个同线群中发现和人类 3 号染色体有同源基因，以虚线分开。因此，小鼠 16 号染色体的基因代表了人类基因组的 7 个同线群。

这两个物种同线区同源性的程度非常显著。小鼠染色体的 731 个基因中，717 个（98%）与人类基因有同源性。这一高度同源性在很大程度上遮掩了这样一个事实：小鼠一条染色体可由 6 个不同的人类染色体上的同源区段和 7 个自身的同线群所代表。染色体在进化中不断变化，但有关基因表达的染色体区及大的同线群内的基因次序改变不大，当染色体损伤和易位时才会发生这种改变。例如，麂科的两个亲缘关系很近的种，经过大量的染色体断裂（或连接，或二者都有），使这两个物种发生了分化，以至于一个只有 3 对染色体，另一个却有 23 对！然而，这两个物种经过杂交后，可以产生健康但不育的后代。

小鼠和人在基因组学上的相似程度，明显地与这两个物种的外表和行为不成比例。如何解释这种矛盾呢？如果不能从基因本身找到答案，那就应该从基因的表达方式上研究。有些答案已被证实。我们知道人类基因常常需要经历大量的选择性剪接，据估计，实际上人类基因的 40%在体内有至少两种不同的剪接方式（第 14 章），这使人类蛋白质组（全部人类蛋白质）比基因组更为复杂。有证据表明人类基因与其近亲黑猩猩的同一组基因在表达模式上有非常大的差异，与小鼠的差异甚至会更大。这可能源于 microRNA 的控制，相对于蛋白质编码基因，microRNA 在小鼠和人类之间的差别似乎更大些。

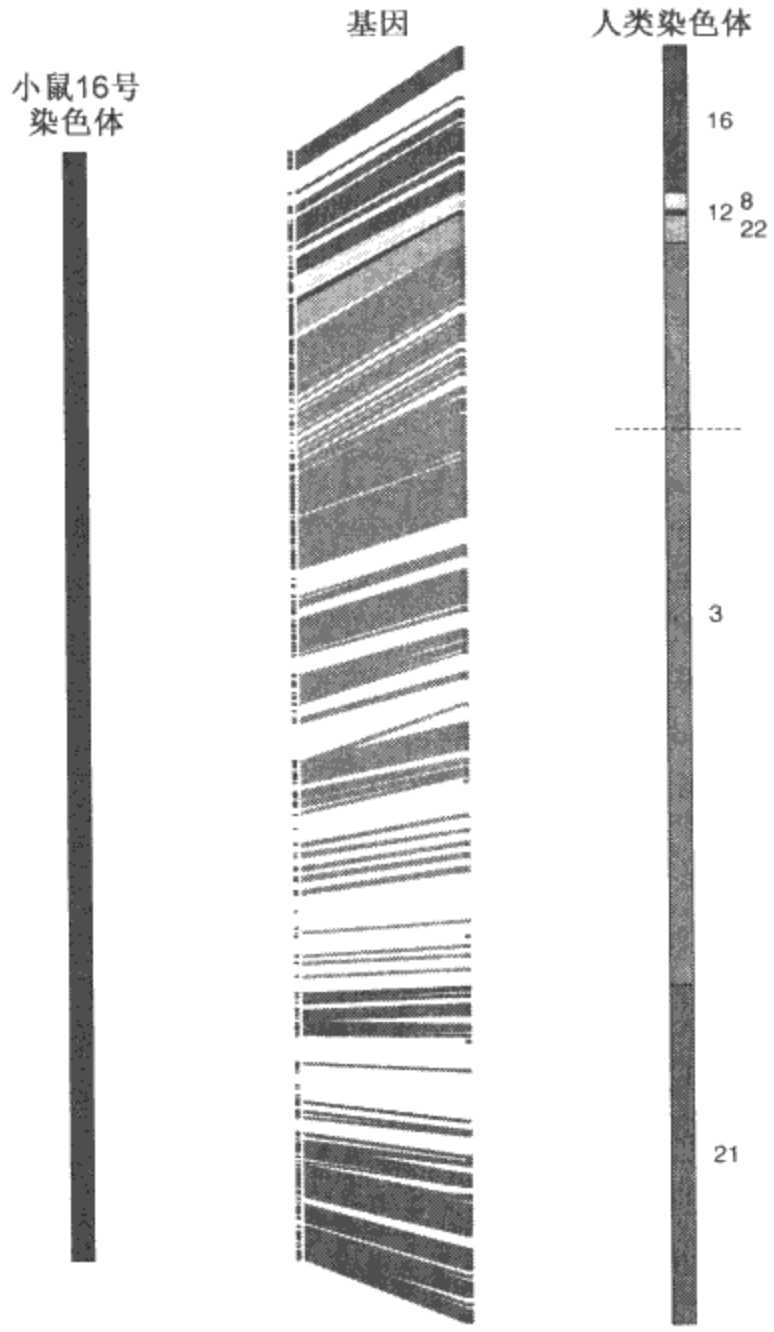


图 24.14　小鼠 16 号染色体与人类基因组的共线保守区域。通过蛋白质水平的分析检测同源基因。小鼠 16 号染色体在左边标出，人类 6 个不同染色体的同线区在右边标出（不同颜色表示人类不同染色体）。小鼠和人的种间同源基因用彩色线连接起来（如中间图所示），以细水平线所示（蓝色，小鼠；红色，人类）。人类 3 号染色体上与小鼠 16 号染色体上的同源基因可在两个明显不同的同线群中找到，以虚线分开。线上方的是人类 3q27～29 染色体区域；线下方的是 3q11.1～13.3 区域。（*Source*：Adapted from Mural et al., *Science* 296 (2002) Fig. 3, p. 1666.）

小结　比较人类和其他脊椎动物基因组，发现了很多相似和不同之处。这些比较还有助于鉴定一些人类基因，而且也有助于找出人类遗传病的致病基因。我们还可以利用亲缘关系相近的物种，如通过研究小鼠基因的表达规律，预测人类相应基因的表达规律。小鼠和人类染色体的详细比较表明，这两个物种之间有很高的相似性。

最小的基因组

到2002年初，已有50余种细菌基因组被测序。其中最小的应属于胞内寄生菌，如支原体、发疹伤寒等的病原体（一种能够引起高山眩晕发热的病原体家族成员之一），以及寄生螺旋菌，如包柔氏螺旋体菌（*Borrelia burgdorferi*，一种引起莱姆关节炎的细菌）。作为最小细菌基因组的记录一直由生殖道支原体（*Mycoplasma genitalium*）保持，仅有530kb。以上分析使遗传学家不禁要问，“能够维持生命的最小的基因组是什么?”

解答该问题的一种方法是比较各细菌基因组，发现其最小共有部分，即它们共同拥有的基因。但是这样产生了一组含有约80个基因的基因组，明显太少而不能维持生命。因此，不同的细菌必然有其不同途径使它们的基因组呈连续性，仅简单地找出这些不同途径相叠的终点是没有用的。

1999年，Craig Venter及同事报道了用另一种方法发现最小基因组的结果。他们利用转座子干扰基因，系统地对生殖道支原体及相关的肺炎支原体（*Mycoplasma pneumoniae*）基因进行突变，然后观察哪种基因是必需的，哪种是非必需的。他们发现在这些生物中，480个蛋白质编码基因中的265～350个是必需的，但其中的111个功能尚属未知，这非常令人吃惊，表明仍然有很多东西要了解，究竟是什么承担了维持生命的重任。

该实验鉴定了**必需基因簇**（essential gene set），即该组基因的缺失是致死性的。必需基因簇不同于**最小基因组**（minimal genome），后者表示真实生物维持生命的基因集合。这种区别源于一个事实：有机体自身可以承受缺失某些基因，但是两个或更多这些基因的同时缺失是致死性的，因此，这些基因不是基本基因组簇，而是最小基因组的一部分。

接下来的任务是找出哪些基因需添加到必需基因簇以产生最小基因组。Venter及同事雄心勃勃地提出：构建携带若干基因的DNA框（DNA cassette），然后将该框转入肺炎菌细胞中，该细胞的自身基因已失活，所以不会混淆结果。他们用不同组合的基因进行实验，直到发现一个含有最少数量，但又能维持生命的基因组合。

这一计划必须处理的难题是如何使这些基因能够在新的、不含有自身任何基因的细胞中正常行使功能，而且Venter及同事也承认该难题可能难以克服。的确，我们可以将一个或几个外源基因转到正常细菌细胞中，并且顺利表达，但全部都是外源基因的情况下又会怎样呢？很可能基因不表达，而仅仅是存在于细胞中。Bernhard Palsson对这一问题是这样表述的：“怎样去激活一个新的基因组呢?”

假设这一技术性问题能够解决，潜在的伦理问题仍然存在：以无生命元件去创造生命道德吗？认识到这些问题，Venter及同事向伦理学家小组递交了他们的计划。1999年，伦理学家小组做出决定认为这一计划不存在严重的伦理问题。但他们确实也看到了安全性问题，并建议公众官员检验Venter及同事将要创造的人造生命形式对环境可能造成的危险性，或能被改造用于生物恐怖或生物战争的可能性。

为什么要构建一个含有最小基因组的生物体呢？从纯科学角度来讲，存在这么一种含有最小基因组的生物很重要，它可以研究为什么需要这些特殊基因，而且还可能有独特的应用价值。事实上，Venter及同事计划补充不同的基因给最小基因组以使细菌能够产生燃料，诸如氢，或者清除代谢废物包括来自于植物的CO_2。

最后，值得注意的是生殖道支原体和其他细胞内寄生菌，由于它们的寄生生活，可以侥幸摆脱小基因组的困扰。它们从寄主中汲取了所需的大部分营养，所以在自身没有所需营养基因的条件下能够生存。事实上，以人为寄主的肺炎菌早已将自身的基因组进化得与维持生命所需最小基因组比较接近，但是科学家们仍然需要将其变得更小，即在严格控制的实验条件下勉强维持生命的最小基因组。

小结 通过每次突变一个基因来观察哪些基因是生命必需的方法，来定义简单生物体的必需基因组是可能的。原则上讲，此方法定义最小基因组——生命所必需的最小基因组也是可能的。很可能最小基因组大于必需基因组。理论上讲，将最小基因组转到缺失自身基因的细胞中并创造新的生命形式也是可能的，这种新的生命形式可以在实验条件下存活和繁殖。添加选择性基因后，这种被改造后的生命形式可以行使多种有用的功能。

生命条形码

分类学家主要从事生物体分类并理解它们的差异和关系，传统的分类学仅仅以生物体的外观和形态特征为基础区分不同的物种。现在已经进入DNA测序时代，他们获得了另一个工具，因为不同的物种有不同的DNA序列以及不同的外观。如果保持恒定的突变率，还可以根据两个生物的DNA序列间差异的程度准确地估算它们之间的遗传距离或两个物种分化的时间。

然而，我们要对数百万物种进行测序，不可能用现在的技术对所有物种的基因组或仅仅一个较大

的基因片段进行测序。但是分类学家可以将研究集中在所研究的物种间有显著差异的一个特定区域。一组叫做生命条形码联盟（Consortium for the Barcode of Life，CBOL）的科学家提出从地球的每一种物种中获得其基因组的一段相对较短的 DNA 序列，该序列相当于条形码。理论上，这使我们能够快速地鉴定任何已知的物种，包括生物恐怖主义的“武器”，而且这也将有助于将新物种放入进化树合适的一个分支上。这项工作可以从已知的 170 万种动物和植物开始并延伸到其余的一千万甚至更多的未知物种（不包括微生物）。

CBOL 的科学家选择了线粒体细胞色素 C 氧化酶亚单位 I（cytochrome c oxidase subunit I，COI）基因中一个 648bp 的特定区段作为条形码，这至少对动物是合适的。所有物种都有这个基因，这个区域在相近物种间表现出一定的差异，但在同一物种的个体间差异很小。例如，这个 648bp 的条形码在不同人之间的差异仅为 1 或 2 个碱基的差异，而人和在进化树中与我们近邻的黑猩猩差异为 60 个碱基。648bp 的序列很容易由自动测序仪获得，由于每个细胞含有 100～10 000 个线粒体 DNA 拷贝，故相对于核 DNA 只有 2 个拷贝而言，线粒体 DNA 更容易纯化。

COI 条形码的一个缺点是植物线粒体 DNA 序列的变异较动物小得多，因此，COI 条形码对植物而言不是很有效。可能包括两个基因在内的一个新区域正在发展为适用于植物的新条形码。

小结 通过建立条形码来识别地球上的所有物种的研究已经开始，第一个“生命条形码”将由每个生物体都存在的线粒体 COI 基因中一段 648bp 的序列组成，此序列足以特异性地鉴定出几乎所有的动物。另一个新的条形码序列是为植物设计的，也许最终可以延伸到微生物。

24.3 基因组学的应用

到目前为止，我们在本章前面的章节中讨论最多的就是寻找和获得基因组序列。现在需要了解获得了一个基因组或一组基因组的序列后究竟有什么用？毫无疑问它有很多用途，所有的用途都可以标记为后基因组（postgenomic），因为其依赖于已经存在的基因组信息。我们已经触及的一个主要应用是利用基因组开展进化研究。另一类应用称为**功能基因组学**（functional genomics），研究基因组的功能和表达。

首先来关注基因组在 RNA 水平上的表达，如果我们在考虑一个生物体在任何给定时间产生的所有转录物，我们称其为**转录组**（transcriptome），类似于“基因组”。这一术语涉及一个生物体的所有基因。在功能基因组学研究中分出一个领域，即同时检测许多基因的 RNA 产物水平，称之为**转录组学**（transcriptomics）。

其次，利用基因组学的知识系统地逐次阻断一个基因的表达，并检测其对某一生物体或生物体某一功能的特定方面的影响，这一类分析称之为**基因组功能谱分析**（genomic functional profiling）。

再次，可以对多个个体的基因组进行比对，以发现它们之间的显著差异。例如，单个核苷酸的不同被称为单核苷酸多态性（single-nucleotide polymorphism，SNP）。有时这些 SNP 与遗传紊乱或没有明显特征疾病（如药物的敏感性）密切相关。但 SNP 并不是人类基因组间唯一的常见差异，遗传学家看的越多，发现的染色体结构变异也越多，如倒位、重复和缺失。而且，至少这些变异中的一些显现重要的作用。例如，在欧洲人中常常发现一个倒位片段，但在非洲人和亚洲人中没有，并且带有这个倒位片段的妇女比不带有这个倒位片段的妇女有更多的孩子，这个倒位似乎提供了某种进化优势。

最后，可以研究基因组蛋白质产物的结构和功能，在一定程度上着重于蛋白质结构的研究，被称为**结构基因组学**（structural genomics），但整体被称为**蛋白质组学**（proteomics），本章的后一节再阐述这个学科。本节主要讨论转录组学、功能基因组学、基因组功能分析和 SNP。

转录组学

要揭示特异组织中某个基因随时间的表达模式，可以进行斑点印迹分析（dot blot analysis），该方法在第 5 章中已讨论过。在标准的点杂交中，将含有要研究基因的单链 DNA 点在膜上形成直径为几毫米的点，然后用标记的 RNA（从所研究的组织中分不同时间提取获得）与膜上的点进行杂交。但是，如果想同时研究特定组织中所有基因随时间的表达模式，情形会怎样呢？原则上讲，可以将同一个细胞中所有 mRNA 所对应的成千上万个单链 DNA 全部点在一张膜上，然后用标记的细胞 RNA 与此膜进行杂交。但如果是这样，超大的膜会产生严重问题。幸运的是，分子生物学家设计了一些新方法来最大限度地缩小膜，而且开发了一些新的技术来分析整个基因组的表达情况。我们先来了解一下 DNA 微阵列和基因芯片，然后再深入到一些更加奇妙的技术领域。

DNA 微阵列和芯片 为了解决膜的大小问题，

分子生物学家利用喷墨打印机技术将极少量的 DNA 点在一个很小面积的芯片上，所以每一个点都非常小。这就允许很多不同的 DNA 点在同一张芯片上，此为 **DNA 微阵列**（DNA microarray）。由 Vivian Cheung 及同事研制的一种机器人系统，可以同时用 12 个平行的微针管操作，每一针管可挤出非常微量的 DNA 溶液（0.25～1.0nL，1/1 000 000 000L）。这些点小到直径仅 100～150μm，而点与点的间距只有 200～250μm。微阵列的结果看起来就像图 24.15 所示。图 24.15 中显示的是在一张普通的显微镜载玻片上只有 7500 个 DNA 点的微阵列，如果点的密度提高，结果会更好。点样后，玻片上的 DNA 被风干，同时在紫外光下使 DNA 与玻片上的硅烷化表层共价交联。

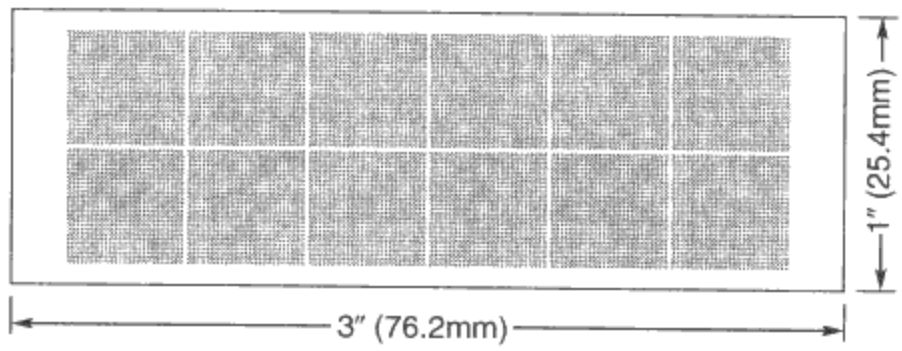


图 24.15　DNA 微阵列示意图。图示标准的 1″×3″显微镜载玻片上的一个 7500 个 DNA 点的微阵列。每个点的直径为 200μm，点间距为 400μm。这并不意味已达到了最高密度。实际上在这样一张玻片上可以放置超过 50 000 个点。（*Source*: Adapted from Cheung, V. G., M. Morley, F. Aguilar, A. Massimi, R. Kucherlapati, and G. Childs, Making and reading microarrays. *Nature Genetics Supplement* Vol. 21（1999）f. 2, p. 17.）

另一个缩小膜尺寸的策略是直接在芯片的表面合成许多寡聚核苷酸。Steven Fodor 及同事在 1991 年创立了这种方法，他们采用计算机芯片制作中所使用的照相平版印刷技术（photolithographic technique），在极小面积和空间里合成短链 DNA（寡聚核苷酸）。这一技术发展到 1999 年（图 24.16），研究人员开始使用附加有人工合成接头的小玻片开始寡聚核苷酸的合成工作，这些接头起初都被光敏基团（能够与光产生反应从而被去除）所封闭。将玻片的某些区域遮蔽，然后对玻片进行光照处理，使未遮光区域的光敏基团被成功去除，随之利用化学合成法将同一核苷酸加在所有去除了光敏基团的接头上。接着，又将另一组点阵进行遮盖，光照去除其他点阵上的光敏基团，加入另一种核苷酸。在两步合成中都未被覆盖的点上，形成二核苷酸。通过重复以上过程，可以在这些点上建立起不同寡核苷酸的微阵列。由此产生的芯片称为**DNA 芯片**（DNA microchip）或**寡核苷酸阵列**（oligonucleotide array）。

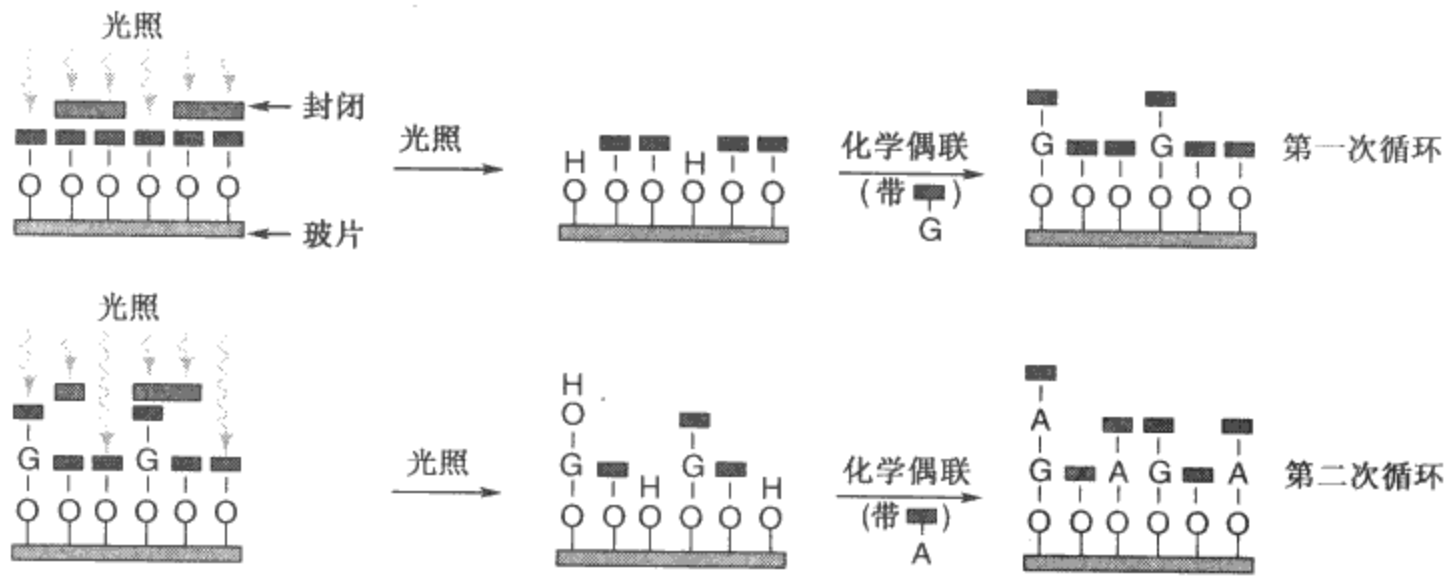


图 24.16　在玻璃基质上不断伸长的寡聚核苷酸链。玻片上包被着反应底物，而这些底物被光敏基团所封闭（红色）。这些封闭剂可以被光照所消除，而其中被遮盖的部分（蓝色）光不能穿透。在第一轮反应中，6 个点中的 4 个被遮蔽起来，只有 2 个可被光照去除封闭剂，一个自带封闭剂的鸟苷酸被加到一个去除封闭的底物上。在第二轮中，3 个点被遮蔽起来，剩余的 3 个暴露在光照下，这包括早已加上了 G 的点。这样，在一个自带封闭剂的腺苷酸被加到 3 个去封闭的点上后，第一个点变成了 G-A 两个核苷酸，第三个和第六个点分别含有一个腺苷酸，第四个是 G，第二个和第五个点在第一轮和第二轮中均被遮蔽，因此没有核苷酸掺入。这样循环下去，每一次遮蔽的位置不同，且加入的核苷酸也不同，最后在每一个点上都会形成各不相同的寡聚核苷酸链。

这一技术如此神奇，它能够将大约 300 000 个寡聚核苷酸点微缩在仅 1.28cm×1.28cm（约 0.5in^2）的面积上。而且这一过程非常高效，能使 4^n 个不同的寡聚核苷酸只需在 $4\times n$ 个循环中完成合成。因此，如果目标是产生由 4 种核苷酸产生的所有 9mer 寡聚核苷酸的组合（4^9，或者 250 000 个不同的寡聚核苷酸），只需要 $4\times 9=36$ 个循环就可以完成。那么，将一个人类基因从基因组中完全无误地鉴别出来，所用的寡聚核苷酸链需要多长？如果掌握了人类基因组的序列将有利于我们精确地回答这一问题。但是，即使我们没有这些信息，仍然可以做一个最小估算。一个有 n 个核苷酸的 DNA 序列在每 4^n 个碱基中出现一次，换句话说，一个 DNA 序列要在长度为 4^n 个碱基的 DNA 中出现一次，其长度必须是 n 个核苷酸。这样我们需要解下列方程求出 n，找出在整个人类基因组 3.5×10^9 bp 的序列中仅出现一次的寡聚核苷酸长度的最小值为

$$4^n=3.5\times 10^9$$

答案是：如果 $n=16$，$4^n>3.5\times 10^9$。因此寡聚核苷酸至少需要 16 个碱基，这需要 $4\times 16=64$ 个循环来建立一个寡聚核苷酸阵列。然而我们已经谈到这是一个最少的估计，最好是比以上所计算的长度长一些，以确保其在人类基因组中只重复一次。

早在人类第一条染色体的序列发表以前，Affymetrix 公司的科学家们就已经设计了识别单个基因的 25mer 的寡聚核苷酸芯片。他们的设计基于当时的已知序列，包括很多数据库中的表达序列标签（EST）。为了增加所制作的芯片的可信度，他们将每个寡核苷酸在一张芯片上进行了多次重复，以使获得的结果可以进行相互对比。

芯片上的寡核苷酸或微阵列上的 cDNA 都可以与标记的细胞 RNA 或相应的 cDNA 进行杂交以检验细胞中哪些基因正在转录。例如，Patrick Brown 及同事利用 DNA 微阵列技术来检测血清对人类细胞 RNA 合成的影响。他们分别从含或不含血清的细胞中分离 RNA，进行反转录，并在此过程中将带有荧光染料的核苷酸掺入 cDNA 进行标记。绿色荧光代表不含血清的人类细胞样品，红色代表含血清的细胞样品。然后将两种标记的 cDNA 混合，与含有 8613 个不同人类基因的 cDNA 微阵列杂交，最后检测荧光信号。图 24.17 显示三次杂交的同一区域的 DNA 微阵列杂交荧光信号，红点表示被血清诱导转录的基因，绿色表示无血清细胞中转录的基因，黄点则表示两种探针同时杂交到同一个点上，因此黄点表示无论有无血清均能够活跃转录的基因。

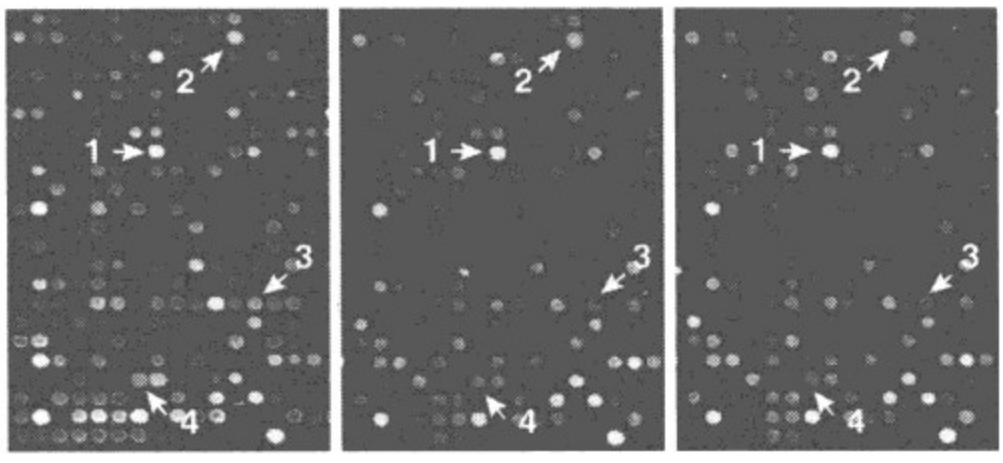


图 24.17　DNA 芯片的应用。Brown 及同事分别从血清刺激和缺乏血清的人类细胞的 RNA 反转录获得 cDNA。他们将缺乏血清的样品标记为绿色荧光；血清刺激的样品为红色荧光。然后同时用两种标记的探针与含有 8600 个人类基因的 DNA 芯片进行杂交。图中显示同一芯片三次不同杂交的结果，展示的区域属于芯片的同一个部位。红点（即点 2 和 4）对应血清存在条件下比较活跃的基因。绿点（点 3）表示缺乏血清时较活跃的基因。黄点（点 1）表示在以上两种情况下的活性基本相同。（*Source*: Lyer, V. R., M. B. Eisen, D. T. Ross, G. Schuler, T. Moore, J. C. Lee et al., The transcriptional program in the response of human fibroblasts to serum. *Science* 283（1 Jan 1999）f. 1, p. 83. Copyright © AAAS.）

微阵列使我们能够在一个更为复杂的体系中检测到基因表达的变化。例如，酵母基因组全序列的获得使分子生物学家能够利用 DNA 芯片在不同条件下对酵母的每一个基因同时进行分析。

在另一个例子中，Kevin White 及同事在 2002 年利用 DNA 芯片对果蝇的全部生活史中的 66 个不同时段的 4028 个基因的表达情况进行了跟踪调查。图 24.18a 显示从 66 个不同发育时段收集的 RNA 用于基因表达分析。我们注意到几乎有一半的时段（30 个）处在胚胎发育时期，这期间也是基因表达

变化最快的时期。实际上，在胚胎发育的早期阶段基因表达十分活跃，每隔半小时就取一次样。以上分析得到了下面几个结论。

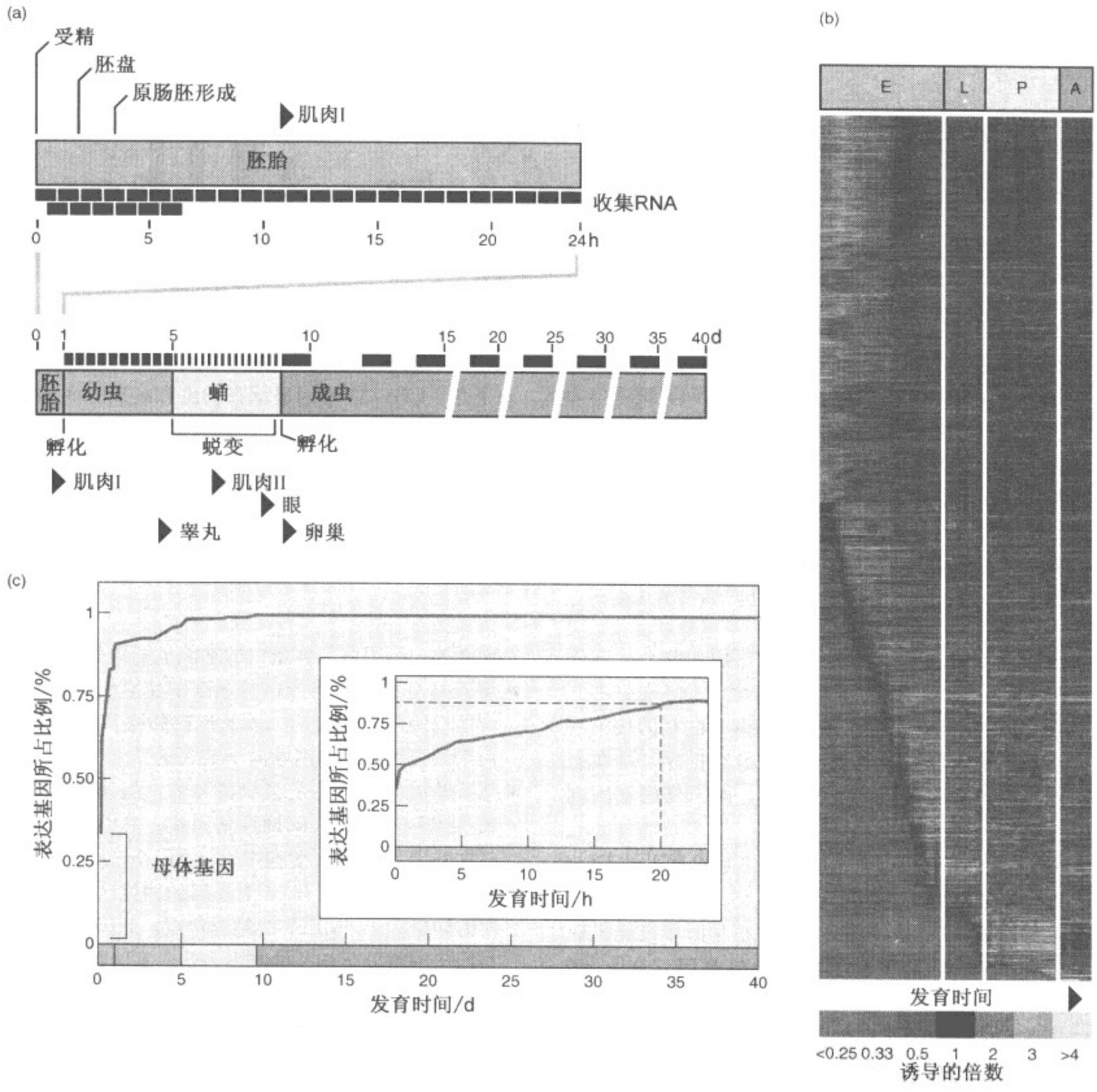


图 24.18　果蝇发育过程中基因的表达模式。（a）收集 RNA 样品的时间分布图。White 及同事在发育过程中按既定的时间以整个虫体为样品收集 RNA。将胚胎时期放大以示所有重叠的取样点。用 oligo（dT）-纤维素层析法纯化了poly（A）$^+$RNA，并在反转录过程中用荧光素标记 cDNA。然后将特定时间标记的 cDNA 探针与微阵列杂交并检测荧光信号。同时以生活周期各发育阶段 RNA（E：胚胎的；L：幼虫的；P：蛹的；A：成虫期的前 40 天）混合物的杂交值为参考标准，与之对照使获得的所有杂交值标准化。（b）基因表达分析。3219 个基因（在果蝇的整个生活周期中表达量的变化达到或超过 4 倍）按照第一次转录产物有显著增加所出现的时间进行排序。发育时期标在上端，所用字母及颜色代表的含义同图（a）。图的底端不同颜色代表表达水平的高低，蓝色为低，黄色为高。（c）图示表达显著增强的基因累计所占比例。值得注意的是，在最早的时间点已有很大比例（约 33%）的基因转录产生了大量的 RNA。这些基因标记为母体基因。插图部分是胚胎阶段前 20h 情况的放大，说明在发育的前一小时大部分转录产物已经出现。

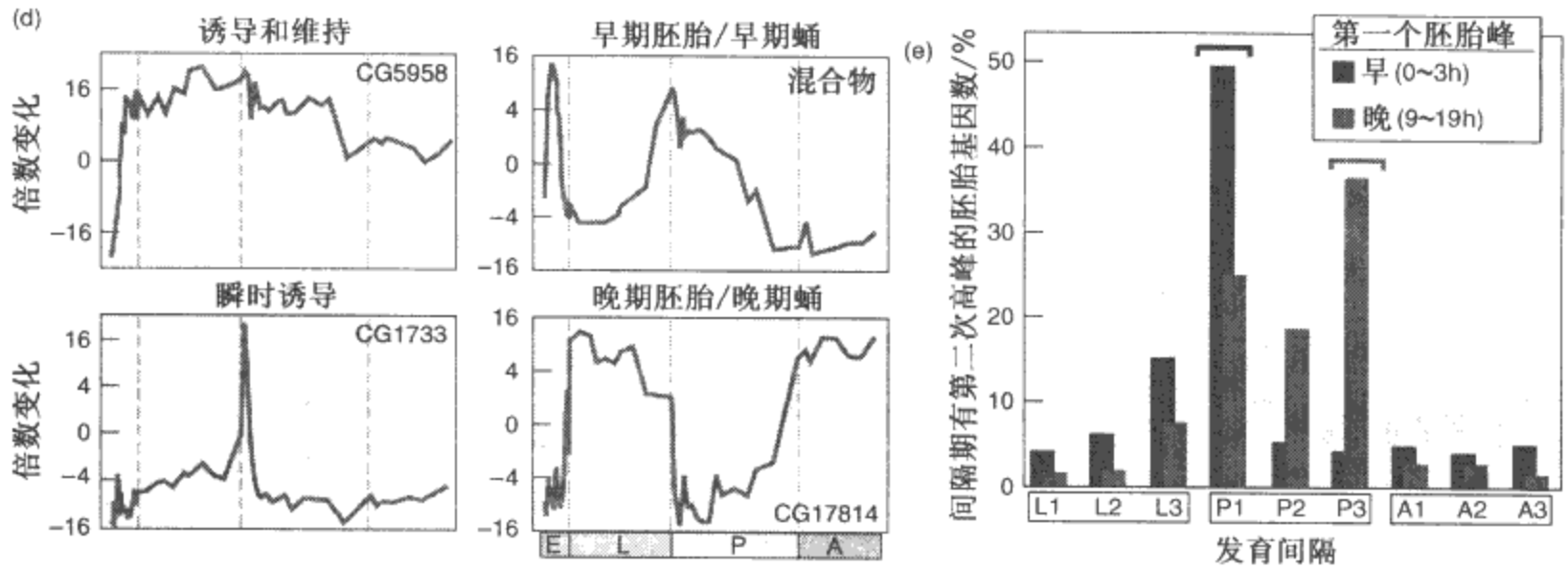


图 24.18　果蝇发育过程中基因的表达模式（续）。（d）4 个基因的表达模式。左上角，CG5958 基因在胚胎早期就被诱导较高表达，并在整个生活周期一直维持高水平。右上角，*Amalgam* 基因表现为胚胎早期诱导，幼虫期降低，在幼虫与蛹期之间又被诱导表达。左下角，CG1733 基因显示在幼虫和蛹期之间的交界处有一个独特的表达高峰。右下角，CG17814 基因在胚胎晚期表达突然增强，一直持续到幼虫期结束，然后在蛹晚期再次被诱导。（e）再诱导模式。在胚胎发育早期（蓝色）或晚期（紫色）表达均显著增强基因所占百分率说明一些基因在发育后期再次被诱导。值得注意的是，在胚胎早期被诱导表达的基因在蛹发育早期也倾向于再次达到高表达（P1，在蓝色柱上加有括弧）；而在胚胎晚期被诱导的基因在蛹发育晚期也将达到高表达（P3，在紫色柱上加有括弧）。（*Source*：Adapted from Arbeitman et al., *Science* 297, 2002. Fig. 1, p. 2271 © 2002 by the AAAS.）

- 很大一部分基因（3219 个）在果蝇的整个生活周期中的表达都经历了较大的变化（4 倍或更多）。图 24.18b 显示的是所有受发育调节的基因按第一次表达显著提高所出现的时间进行排序。排在表最上面的基因是最早被诱导的，而下部的基因则是后期诱导的。
- 超过 88%的发育调控基因在胚胎阶段的末期前 20h 之内表现非常活跃（图 24.18c）。
- 大约有 33%的发育调控基因在很早阶段已经出现（图 24.18c）。这些基因代表**母体基因**（maternal gene），是在母体的卵子发生过程中表达的基因。因此，成熟的卵母细胞已开始转录这些基因，使对应的 mRNA 早就存在，并且在受精开始的第一时间里就可以翻译成蛋白质。
- 正如在图 24.18d 中所描述的，某些基因在整个生活史中均维持一定的表达水平，不受其他基因表达增加或减弱的影响。这一点在图 24.18e 中做了进一步的阐述，在胚胎早期达到表达高峰的基因也倾向于在蛹发育早期再次达到表达高峰；而在胚胎晚期达到表达峰的基因也会在蛹发育晚期再次达到高峰。另外一个与此类似的现象在此未做详细阐述，即在幼虫发育时期高表达的基因在成虫期也倾向高表达。
- 编码特定超级大分子复合体组分的基因倾向于其表达。因此，编码核糖体蛋白的基因倾向于被协同调控，编码线粒体蛋白的基因也同样如此。
- 编码具有相关功能蛋白的基因即使它们不形成复合体也倾向于共表达，因此编码转录因子或细胞周期调控因子的基因倾向于共表达。
- 某些基因的共表达是组织特异的。例如，一个由 23 个基因组成的共同调控基因簇，其中包括 8 个已知基因是在肌肉细胞中表达的。经过进一步的检测，此基因簇中有 15 个基因的调控区域含有成对的转录因子 dMEF2 的结合位点，此因子能激活与肌肉细胞分化相关的基因表达。基因簇中另外 7 个的功能未知，但其中的 6 个具有 dMEF2 的结合位点并在分化中的肌肉细胞中表达。由此，White 及同事推断这 6 个基因在肌肉的分化过程中发挥作用。这一点很重要，因为仅仅根据基因的序列来确定其功能通常是很难的，而基因定时和定位表达研究所获得的相关线索将有助于确定基因的功能。的确，这些线索使得 White 及同事确定了他们所分析的 53%基因的功能。

小结　功能基因组学是对大批量基因表达的研究。该领域的一个分支是转录组学，即转录组的研究。转录组是一个生物体在任一给定时间产生的所有转录物。转录组学研究的方法之一是创建 DNA 微阵列或芯片，其上分别载有成千上万的 cDNA 或寡聚核苷酸，然后与标记的 RNA 或相应的 cDNA 杂交。每个点杂交后的荧光信号强

度揭示出对应基因的表达水平。根据微阵列获得的信息，我们可以同时详查许多基因的表达模式（时间上和空间上的）。基因在时间和空间上成簇表达模式揭示了这些基因产物在某些过程中的协同作用。如果一个未知基因与一个或多个已研究清楚的基因同时表达，则将为我们推测未知基因的功能提供有关线索。

基因表达系列分析 1995 年，Victor Velculescu 和 Kenneth Kinzler 及同事建立了一种分析细胞中基因表达的数量或程度的新方法，称之为**基因表达系列分析**（serial analysis of gene expression，**SAGE**）。SAGE 的基本策略是合成对应于一个细胞中所有 mRNA 的短 cDNA 或**标签**（tag），然后将这些标签连接在一起克隆并测序，以此来了解标签的特性，及细胞中表达基因的特性以及每个基因表达的程度。

图 24.19 说明了 Velculescu 及同事如何开展此项研究。首先，他们用带有生物素的 oligo（dT）引物对人胰腺组织中的 mRNA 进行反转录，获得双链 cDNA。由于短标签（图 24.19 中以 9bp 为例）能局限在一个小区域特异性地识别 cDNA（这对此项研究非常重要），为了实现缩短 cDNA 长度以获得短标签并继而将标签连接起来进行测序的目的，Velculescu 及同事利用锚定酶（anchoring enzyme，AE）对生物素化的 cDNA 进行切割，从 3′端切下一个片段。他们选择 *Nla*III 作为锚定酶，它可识别 4 个碱基的限制位点，从而产生平均 250bp 长短的片段，将这些生物素化的 3′片段连接到能结合链霉生物素的亲和素珠子上。

接着，将附着在珠子上的 cDNA 片段分成两部分，一部分连接到接头 Y 上，另一部分连接到接头 Z 上。两种接头上都含有 IIS 型限制性内切核酸酶［**标签酶**（TE）］的识别位点，该酶能够识别此位点下游 20bp 处并进行切割。利用标签酶 *Fok*I 切割 cDNA 片段的结果是得到一组短片段，每一个片段分别含有接头 Y 或 Z，接着是 4bp 的锚定酶识别位点，然后是从 cDNA 来的 9bp 序列。这个来自 cDNA 的 9bp 片段即是标签。如果标签酶切割后留下突出端，可将它补平后备用。

Velculescu 及同事接下来的任务是沿着既定的 DNA 顺序将标签连接起来，这样就可以辨认一个标签序列结束和下一个标签序列开始的位点。为了实现这一目的，他们通过平末端将标记了的片段连接在一起形成中间为两个毗邻的标签（形成双标签）而两端为接头的结构。在接头中含有与一对引物互补的序列，故 PCR 能够扩增整个片段。Velculescu 及同事用锚定酶切割 PCR 扩增产物后，将这些限制性片段连接起来，并进行克隆。由于在每一个标签序列的一端都带有一个锚定酶识别位点，因此二联体很容易识别。一个克隆中至少含有 10 个标签（有些超过 50 个），一次 PCR 扩增即可以对它们进行分析并测序。如果分析的克隆足够多，就能够判断基因表达的程度，如反复出现的标签就说明对应的基因表达非常活跃。

Velculescu 及同事利用 SAGE 分析人类胰脏中基因的表达，结果与预测相符。其中两种最常见的标签（GAGCACACC 和 TTCTGTGTG）所对应的基因分别是羧肽酶原 A1 和胰蛋白酶原 2，这两种在胰脏中大量表达的酶原经过成熟加工，可以对小肠中的蛋白质进行消化。在大量的标签中还鉴定出许多其他熟悉的胰脏基因；然而还有大量的标签序列找不到基因与其匹配，其身份不得而知。当数据库扩大包含全部人类基因时，所有的标签序列都应该有对应的基因，即使某些基因的功能还不清楚。

小结 SAGE 能够测定组织中哪些基因在表达以及表达的程度。特定基因的特征短标签产生于 cDNA 并通过接头连接在一起。对连接的标签进行测序，即可确定哪些基因表达、表达量如何。

全染色体转录作图（whole chromosome transcriptional mapping） 转录组学研究已经成熟，对转录物制图足以精确确定其在染色体中的位置，这类研究被称为**转录作图**（transcriptional mapping），目前正用于阐明前面提到的一个令人费解的问题，即编码蛋白质基因在人类中的数目并不比进化地位很低的蛔虫多，如何才能将这一现实与人类的极度复杂性联系在一起呢？逐渐浮现的答案是蛋白质编码基因的转录物仅仅为人类转录组的一部分。而随着对此问题探究的深入，发现人类转录组比我们预想的更为复杂。

如果我们仅仅考虑蛋白质编码基因的外显子，可以推测只有整个人类基因组的 1%～2%被表达为细胞质 RNA。然而，早在 2002 年，Thomas Gingeras 及同事就用微阵列分析研究人类 21 号和 22 号染色体的表达，发现在人类细胞的细胞质中多聚腺苷酸化［poly（A）$^+$］RNA 的数量比根据这两条染色体编码蛋白质外显子推算的转录物的数量扩大了一个数量级。这些意外增多的转录物被称为**未知功能转录物**（transcript unknown function，**TUF**），在此类分析中被检测到的所有转录区（包括外显子区和 TUF 类似区）被称为**转录片段**（transcribed fragment）。

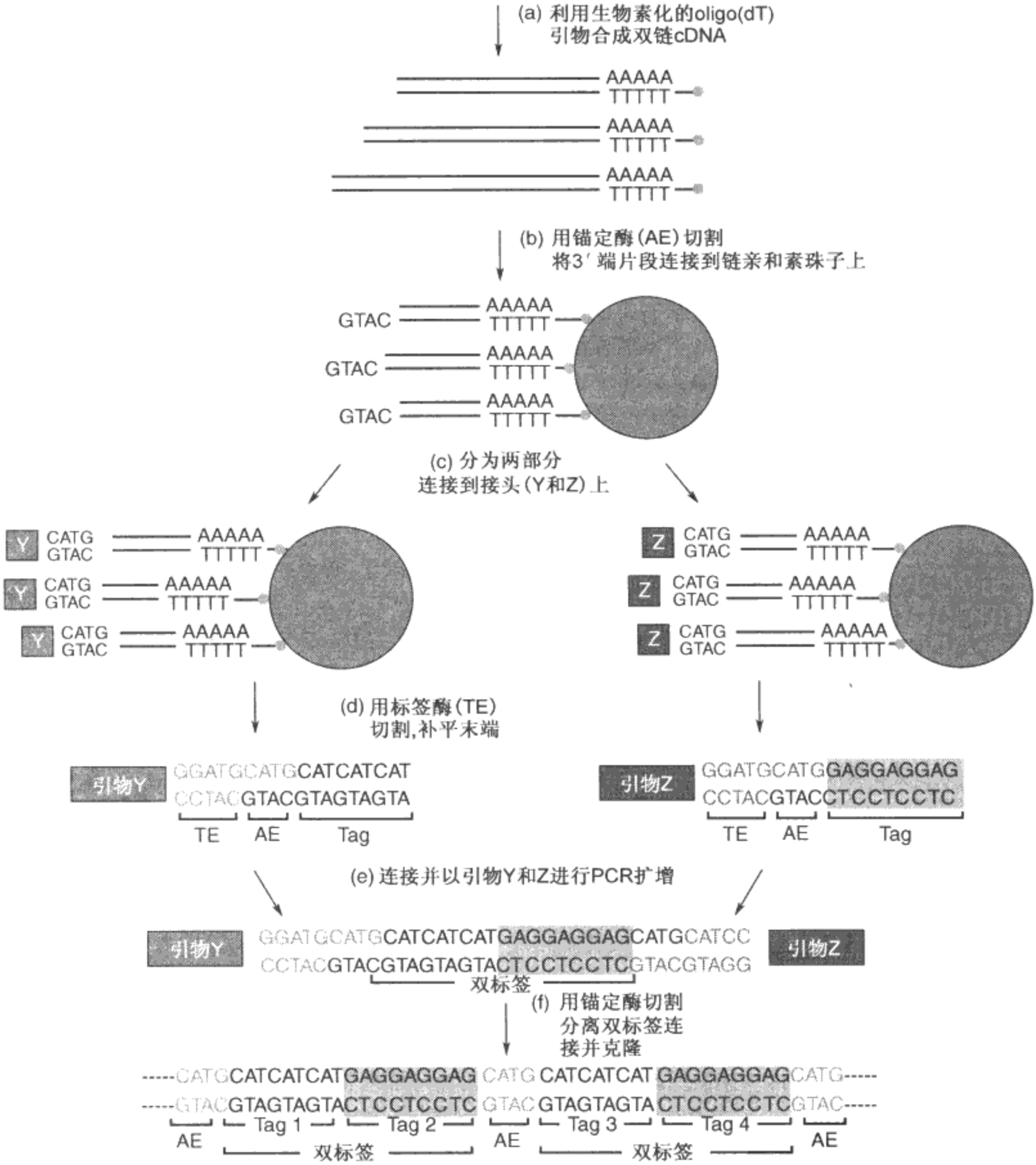


图 24.19　基因表达系列分析（SAGE）。（a）双链 cDNA 由细胞 mRNA 形成，用生物素化的 oligo（dT）引发第一条 cDNA 合成，橘红色球代表生物素。（b）生物素化的 cDNA 用锚定酶（这里是 *Nla* III）切割，生物素化的 3′端片段被连接到链亲和素珠子上（蓝色）。（c）连接在珠子上的片段被分为两部分：一部分与接头 Y（蓝色）连接，另一部分与接头 Z（红色）连接。（d）与接头连接的片段被标签酶（TE）切割，如需要，将末端补平以产生平端。这里用的标签酶是 *Fok* I，经它酶切剩下 9bp 的标签连在接头上。Y 接头上的标签（由任意序列 CATCATCAT 表示）与其互补序列用黄色加亮；Z 接头上的标签（由任意序列 GAGGAGGAG 表示）与其互补序列用紫色加亮。（e）含有标签的片段靠平端连在一起，并由能与接头中的引物 Y 和引物 Z 区杂交的引物进行 PCR 扩增。只有两个标签以尾尾相连形成的片段才能被 PCR 扩增。（f）扩增的包含双标签的片段在锚定酶的作用下产生带黏性末端的双标签序列。这一序列首尾相连形成串联体，被克隆到载体中。双标签串联体的一部分显示在图中，带 4 个碱基识别位点的锚定酶用绿色表示。这 4 个碱基的位点将每一个双标签序列分开，因此很容易辨认。然后测定克隆的序列，由此检测标签代表的序列及其数量。这告诉我们哪些基因表达了，以及它们的活跃程度。（*Source*：Adapted from Velculescu，V. E.，L. Zhang，B. Vogelstein and K. W. Kinsler，Serial analysis of gene expression. *Science* 270：484，1995.）

另外，据报道人类细胞和仓鼠细胞中约 2/3 的转录物是非多聚腺苷酸化的［poly（A）$^-$］，这些 poly（A）$^-$转录物对应于基因组另一个相当大的部分，这个区域的大小还不知道，但很明显会非常大。综上所述，这些发现提示编码蛋白质的外显子仅占全基因组序列产生的细胞质 RNA 的一小部分。

为进一步探索这一奇妙的推断，Gingeras 及同事用高密度寡聚核苷酸分析（寡聚核苷酸为 25mer，平均间隔仅 5bp，因此平均有 20bp 的重叠）。为什么用如此高的密度呢？因为这一方面可以检测短的外显子；另一方面用相互重叠的寡聚核苷酸进行杂交，可以更有把握地确信检测到的转录区域确实发生转录了。用于分析的核苷酸来自 10 个人类染色体（6、7、13、14、19、20、21、22、X 和 Y），占人类基因组全长的 30%，Gingeras 及同事检测了代表 8 种不同人类细胞系细胞质 poly（A）$^+$RNA 的 cDNA，以及代表单个细胞系（HepG2）细胞质和细胞核 poly（A）$^+$RNA 和 poly（A）$^-$RNA 的 cDNA。在以上所有分析中，与假基因或 DNA 重复序列重叠的转录片段不在考虑之列。

对每一细胞系而言，在多于 7.4 千万探针对（双链）中约 9%与 poly（A）$^+$RNA 的 cDNA 发生杂交，应用 1/8 原则（“1 of 8” rule），即一探针对必须与 8 个细胞系中的一个细胞系的 cDNA 发生杂交，这时阳性探针的百分比升至 16.5%，这就是“1/8 制图”。在每个细胞系的 10 条染色体中平均 4.9%的核苷酸表达为细胞质 RNA。而在 1/8 制图中此百分比提高到 10.1%。这些发现提示在一个细胞系中这 10 条人类染色体所有序列的至少 10.1%表达为细胞质的 poly（A）$^+$RNA。而 4.9%和 10.1%的差异表明有相当多的细胞系特异性转录存在。

图 24.20 显示了在细胞质 10 条染色体中每条染色体的 poly（A）$^+$转录物构成比例。需要说明的是，来源于基因间隔区和内含子的转录产物尚未注解，这些区域占了 10 条染色体所有转录物（中间的饼图）的多数（57%）。有注解的转录物与分别注释为外显子、mRNA 和 EST 的区域相一致。

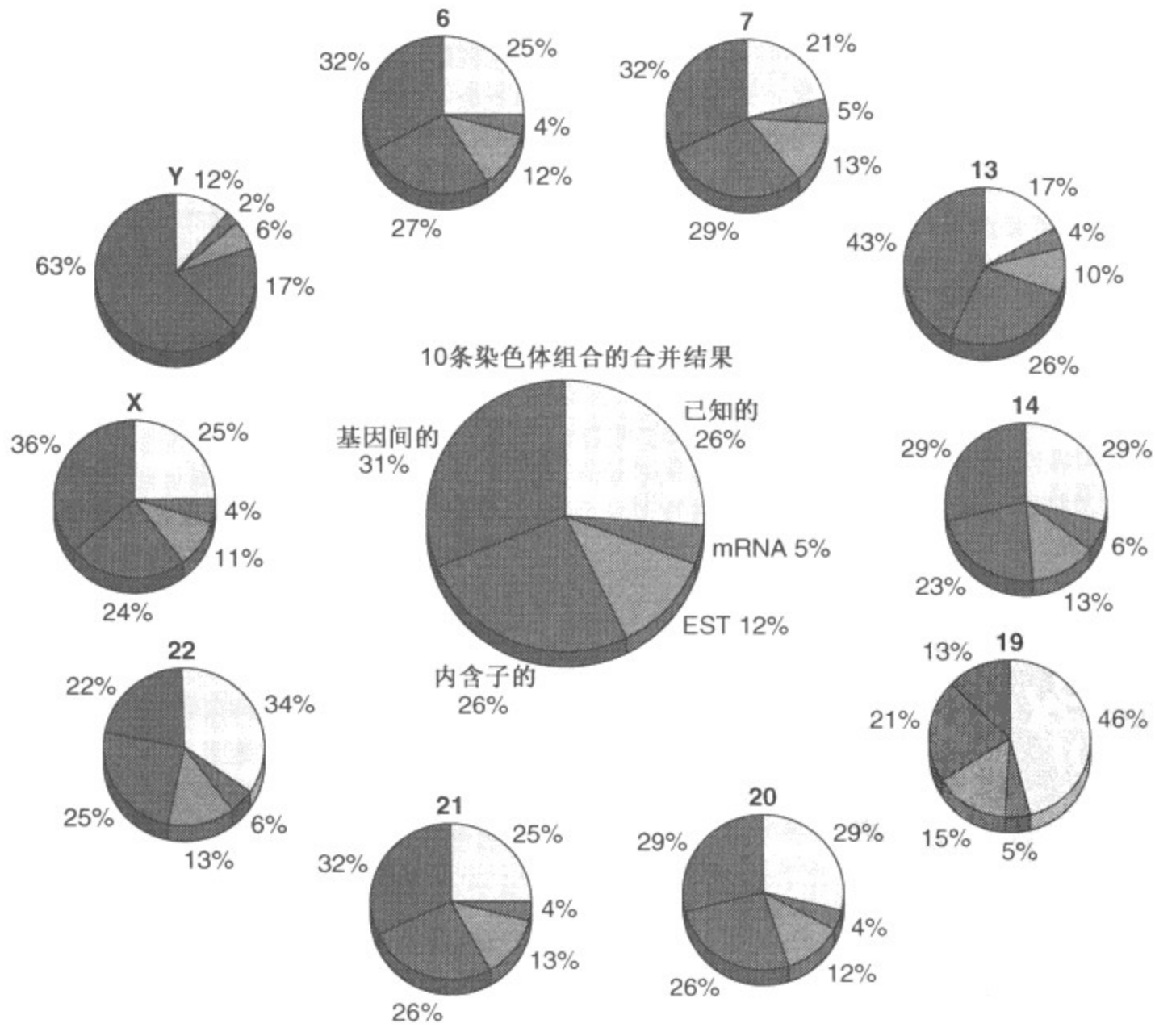


图 24.20　10 条人类染色体的转录图谱。每个饼图中的楔块代表在细胞质 poly（A）$^+$转录物的“1/8 图谱”中不同种类序列所占的百分比。每个小饼图代表（通过黑体字识别）一条染色体，中间的大饼图代表 10 条染色体汇总在一起的情况。序列种类在集合饼图中标出，且同一种颜色适用于整张图。基因间序列和内含子归为尚未注解的序列。外显子、mRNA 和 EST 为注解序列。

poly（A）$^-$转录物的情况如何呢？为了对此进行分析，Gingeras 及同事将研究集中在单个细胞系，HepG2。他们在这些细胞的细胞核和细胞质中寻找 poly（A）$^+$转录物、poly（A）$^-$转录物和双转录物（bimorphic transcript）［双转录物起初为 poly（A）$^+$转录物，随后 poly（A）尾丢失］。他们发现 10 条染色体中至少 15.4%的核苷酸代表了这些种类的转录物之一［其中近一半是 poly（A）$^-$转录物］。因此，对应于稳定转录物的基因组约 10 倍于仅根据外显子预计的基因组。当然，多数人类基因的主要部分是内含子，所以这一结果初听起来并不令人吃惊。但如果剪除的内含子没有功能，可以预测将被迅速降解，并且也可能在稳定的核 RNA 中占非常大的比例。该研究的另一个结论是人类转录物近一半是重叠的，有两种重叠方式，即在同一条链上或在对侧链上。当然，在对侧链上的重叠将呈现为正义链/反义链对，这将引起 RNAi 反应。实际上，这些互补的转录物是依赖 RNA 的 RNA 聚合酶的产物，在 RNAi 过程中能够扩增 siRNA。如果是这样，这可能代表一种基因表达调控的机制。

这些研究表明存在大量细胞质 poly（A）$^+$转录物和非外显子区的 poly（A）$^-$转录物，这有助于理解生物间的差异。虽然人和黑猩猩的外显子非常相似，但非外显子区已经偏离很远，而这些区域的转录可能会产生一些显而易见的差别。

小结 高密度全染色体转录作图研究证明，细胞质 poly（A）$^+$RNA 的大部分序列源于 10 条人类染色体非外显子区。而且这 10 条染色体的近一半转录产物是 poly（A）$^-$转录物。总的来说，这些结果表明 10 条染色体在细胞质和细胞核内的大部分稳定的转录物源于外显子以外的区域。这有助于解释物种间的巨大差异，如人和黑猩猩，他们的外显子几乎是相同的。

基因组功能谱

基因组功能分析的最终目的是确定生物体在生命周期的各个阶段所有基因的表达模式。即使对简单的原核生物，这也是一项令人望而却步的任务。迄今，对于每种生物而言，这一难题也只是逐渐将各个研究小组的结果组合在一起以求破解。让我们先来考虑攻克这一难题的常规技术。

缺失分析（deletion analysis） 一旦基因组中所有基因都被确定，我们就可以去除其中的一个基因来研究生物体发生的变化。当然，对人类而言这种实验从伦理上是不可行的，但至少在原理上可以在其他脊椎动物上实践，只要这一物种的基因组序列已经测得。对脊椎动物如此之大的基因组进行这种分析，其中逻辑方面的问题将延迟工作的进展，但人们已经以这种方式描绘了酵母基因组功能图谱。

2002 年，Ronald Davis 率领的一个庞大的研究小组报道，他们研制出了一套酵母突变体，在每个突变体中一个基因被一个抗生素抗性基因所置换，抗生素抗性基因一侧为 20mer 的因置换基因不同而不同的序列，由此而使每个置换基因获得一个“分子条形码”（molecular barcode），从而可以被特异地识别。他们总共置换了酿酒酵母中 96%的有注释的 ORF。紧接着，他们检测了这些突变体在 6 种不同条件（高盐、山梨醇、半乳糖、pH8、基本培养基和制霉菌素）下混合培养的生长能力。同时将每种条件下获得的 RNA 与 DNA 微阵列进行杂交以检测基因表达情况。

为了绘制酵母**基因组功能图谱**（genomic functional profiling），Davis 及同事将 5916 个不同的突变体混合培养在上述的不同培养基上，并在不同时间段收集细胞，通过与一个包含条形码互补序列的寡聚核苷酸微阵列杂交来检测每个条形码。如果一个基因在克服给定条件（如存在半乳糖）方面有重要作用，则施加该条件时缺乏这个基因的突变体将从混合培养中迅速消失。实际上，突变体消失的速率与所缺失基因在克服给定条件中的重要性相关。

当研究者应用这一方法描述酵母突变体对半乳糖存在的反应时，发现了几个通过多年研究才获知的与酵母半乳糖代谢相关的基因。同时还发现了 10 个新基因是尚未提及与半乳糖代谢相关的。对野生型酵母和 11 个突变体（以鉴定在半乳糖代谢中起重要作用）的生长特性进行了单独测定，结果呈现于图 24.21 中。正如预测的那样，所有的 11 个突变体在半乳糖培养基上的生长速度都慢于野生型，它们的生长速率为野生型生长速率的 44%～91%。

小结 有几种方法可用于基因组功能图谱分析。在一种叫做缺失分析的突变分析中，突变体创建是利用一个抗生素基因一次替换一个基因，并在一侧连接一段寡聚核苷酸作为分子条形码以识别突变体。然后，将这些突变体混合培养在各种不同的条件下以观察各种突变体消失的速度，从而获得功能谱。

RNAi 分析（RNAi analysis） 通过基因突变“敲除”基因非常费力，迄今也只在酵母中从全基因组水平完成了此项研究，但对一些更为复杂的生物体应该选择更为简单的方法，如 RNA 干涉（RNAi，第 16 章）来“封闭”基因。秀丽隐杆线虫尤为适合开展此项工作，RNAi 甚至可以影响其后代；线虫

进行单性生殖，其繁殖只需要父母中的一方；它包含不到 1000 个细胞，并且其全基因组已完成测序。显而易见，该生物是通过 RNAi 描述基因组功能轮廓图的靶生物。

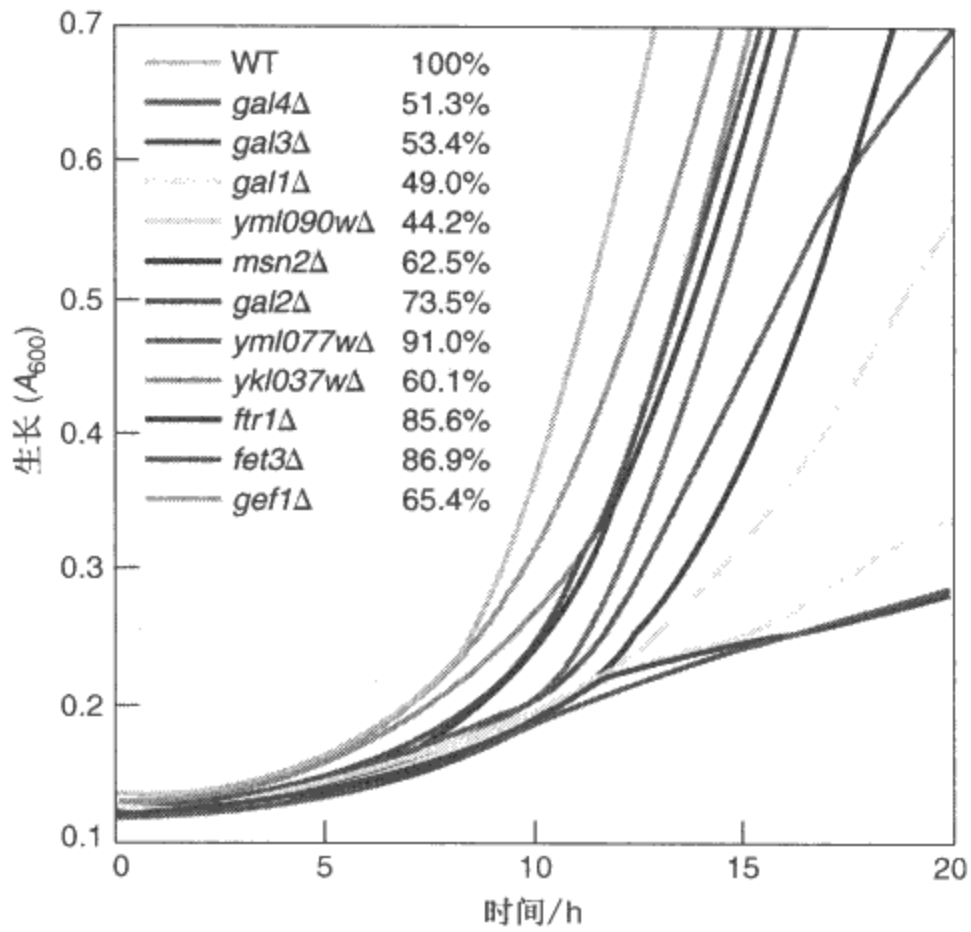


图 24.21　通过基因组功能分析检测出的各种半乳糖缺陷突变体的生长曲线。Davis 及同事在以半乳糖做碳源的培养基上对野生型和 11 个突变的酵母细胞的生长分别进行了测试。11 个突变型菌株被混合后观察其生长速度。A_{600}指混浊度，标志着酵母的生长量和生长速率。（*Source*：Adapted from Giaever，G.，A. M. Chu，L. Ni，C. Connelly，L. Riles，S. Veronneau et al.，Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418，2002，p. 388，f. 2.）

Birte Sönnichsen 及同事已用该技术失活了秀丽隐杆线虫 19 075 个基因，占其全部基因的 98%，并观察这些突变对早期胚胎发生（受精后最初两个细胞的分裂）的影响。他们将 25bp 的双链 RNA 注入虫体并通过时差显微镜分析追踪其后代最初两个细胞的分裂，他们还检查了两细胞阶段之后胚胎的生存能力以及在幼虫和成虫阶段整个虫体表型的改变。

研究者发现 1668 个基因用 RNAi 失活后，获得可检测到的表型缺陷。其中 661 个基因的失活在两细胞分裂阶段造成繁殖能力的缺陷；其余在发育后期阶段造成缺陷（图 24.22；在胚胎发育早期引起缺陷的 661 个基因的有效失活也会产生胚胎性死亡是不足为怪的）。

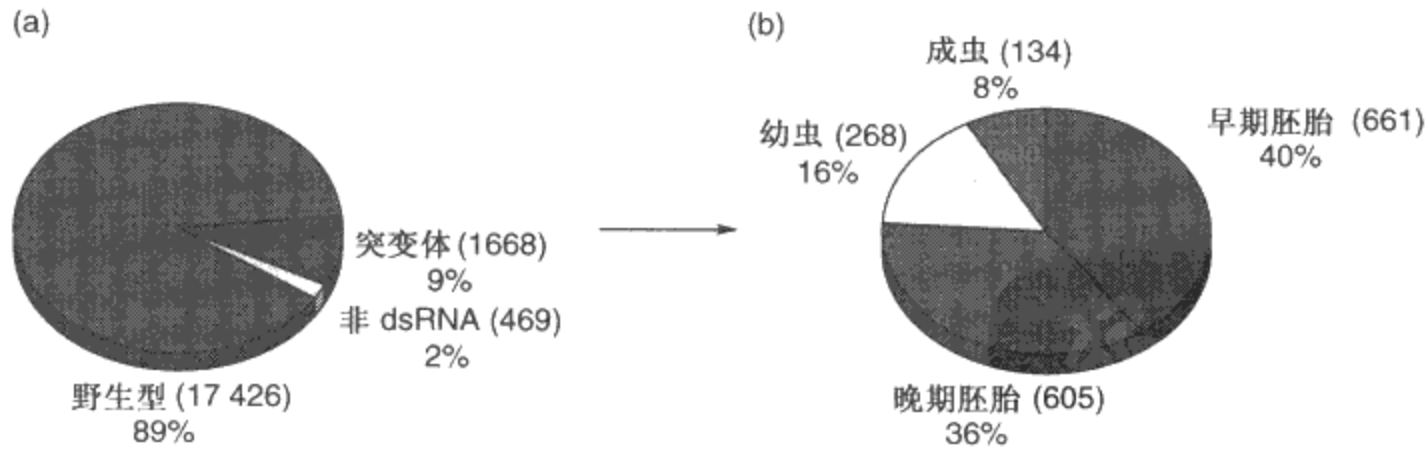


图 24.22　利用 RNAi 获得秀丽隐杆线虫基因组功能谱表型的分布。（a）初筛，Birte Sönnichsen 及同事用 dsRNA 靶向 19 075 个基因。其中 17 426 个（野生型，蓝色）在筛查中没有呈现表型改变，1668 个（突变型，红色）导致表型改变。实验中有 469 个（非 dsDNA，黄色）并不能靶向任何基因。（b）突变表型的分布，Birte Sönnichsen 及同事继而从失活的 1668 个基因（突变表型）中筛选导致发育阶段缺陷的基因。其中 661 个基因（红色）导致早期胚胎（两细胞分裂阶段）缺陷。（*Source*：Adapted from sönnichsen et al.，Full-genome RNAi profiling of early embryogenesis in Caenor habditis elegans. *Nature*. Vol. 434 (2005) f. 2，P. 465.）

RNAi 存在的一个问题是有时不能失活基因（假阴性），所以就很难解释这些阴性结果。Birte Sönnichsen 及同事在检查他们的研究进程时，估计之前已报道的能影响最初两个细胞分裂的基因有 65

个，其中 62 个（95%）在 RNAi 分析中检测到。对在第一次分析中未检测到的 4 个基因再次进行 RNAi 分析，其中 3 个在第二次分析中检测到。

其实只有突变产生明显的表型才能被检测到，所以突变策略同样也产生假阴性。因此，研究者对其研究进程进行了另一项检查，即将他们的数据与其他针对早期胚胎发生的 RNAi 分析进行比较。结果发现，他们已检测到其他研究者所发现基因的 75%。因此，Birte Sönnichsen 及同事谨慎地提出他们的 RNAi 分析检测到 75%～90%的在胚胎发生早期发挥作用的基因。

随后，研究者根据这 661 个基因特异性的表型对它们进行归类，发现这些基因近半数（326 个）从本质上造成胚胎发生的缺陷，而其他基因（335 个）只是影响了一般细胞的代谢，而这在保持胚胎存活足够长以发生两次分裂是必需的。通过对这些特异性缺陷的仔细注解，研究者将前 326 个基因归为胚胎发生中 23 个方面的缺陷，如纺锤体的组装（9 个基因）和姊妹染色体的分离（64 个基因）。

小结 利用 RNAi 失活基因可对复杂的生物体进行基因组功能分析。该方法的一个应用实例是针对参与秀丽隐杆线虫早期胚胎发生基因的 RNAi 分析，结果鉴定出 661 个重要基因，其中 326 个直接参与胚胎发生。

组织特异性功能谱（tissue-specific functional profiling） 基因组功能分析的另一个策略是通过突变或其他手段使基因失活，并观察该基因的组织特异性。Lee Lim 及同事用两种 miRNA 敲低体外培养的人类细胞（HeLa）中基因的表达，继而分析表达明显减少的基因的特性。miR-124 是一个在脑组织中显著表达的 miRNA，能敲低一些基因的表达，并导致这些基因在脑组织中低表达；miR-1 是一个在肌肉组织中显著表达的 miRNA，能敲低一些基因的表达，并导致这些基因在肌肉组织中低表达。换句话说，这两种 miRNA 使 HeLa 细胞中一些基因的表达发生了改变，这正是在分别富含两种 miRNA 的组织中看到的情形。如果在体内两种 miRNA 下调了相同的一些基因，则上述结果正是我们希望看到的。

此研究的另一个显著特点是降低靶 mRNA 的浓度。即使这样，如我们在 16 章中所了解的，动物 miRNA 通常影响 mRNA 的翻译而不是它的浓度。因此，Lim 及同事向 HeLa 细胞中导入双链miRNA，并用微阵列分析处理细胞中纯化的 mRNA 的水平，结果表明每种 miRNA 导致 100 或更多这种 mRNA 浓度的明显降低。

以下为 Lim 及同事如何开展他们的分析。先来看 miR-124，他们首先利用之前全基因组研究获得的数据，绘制了 46 种组织 10 000 个人类基因表达水平的柱状图。图 24.23a 中的柱状图包括了大脑皮层中基因表达的数据，每个柱代表大脑皮层中以某一水平表达的基因数量，最左边的柱代表在这一组织比在任何其他组织表达都高的基因，而最右边的柱代表在这一组织比在任何其他组织表达都低的基因，其他的柱代表中等水平表达的基因，即在大脑皮层中从高表达到低表达。这张图代表了全部的 10 000 基因，而随机的一组基因可产生类似的柱，并作为分析的背景。

图 24.23b 中的柱形图包含在 miR-124 处理的 HeLa 细胞中表达量明显减少的一组基因，而不是背景图，我们看到柱形图明显地偏向在大脑皮层中低表达的基因，请注意右边的柱具有显著的优势，$P<0.001$，实际上 P 值的数量级为 10^{-12}。

接下来，Lim 及同事将 miR-124 影响作用的分析扩大到所有 46 种组织，并对 P 值的对数作图（图 24.23c）。将 P 值小于 0.001 作为显著性差异的阈值，分析的结果表明只有脑组织的 P 值与背景有显著性差异。在 miR-1 影响作用的分析中，Lim 及同事发现只有肌肉组织的 P 值与背景有显著的差异。因此，由 miR-124 诱导的 HeLa 细胞基因表达消沉的模式与仅在脑细胞中呈现的基因低表达模式是一致的。同样，miR-1 诱导的 HeLa 细胞基因表达消沉的模式与仅在肌肉细胞中呈现的基因低表达模式是一致的。

再次提醒大家注意，这些研究是用微阵列检测 mRNA 的水平，因此 miRNA 可能会影响特定 mRNA的稳定水平，即可能破坏 mRNA 的稳定性。如果真是这样，则可以看到 miRNA 和不稳定 mRNA 之间互补的证据。互补区最可能在 mRNA 的 3′-UTR，因为曾发现这类典型的互补存在于此。

因此，Lim 及同事将这两个 miRNA 的序列与表达水平明显降低的 mRNA 的 3′-UTR 序列进行比对，他们用称为 MEME 的“基序发现工具”（motif discovery tool）进行匹配分析，获得了令人注目的结果，所有被 miR-1 下调的 mRNA 中，88%含有一串至少 6 个碱基的保守序列 CAUUCC，此序列与 miR-1 的一串碱基互补。且所有被 miR-124 下调的 mRNA 中，76%含有一串至少 6 个碱基的保守序列 GUGCCU，此序列与 miR-124 的一串碱基互补。这是 miRNA 确实与其靶 mRNA 的 3′-UTR 相互作用并可能破坏 mRNA 的稳定性的强有力证据。

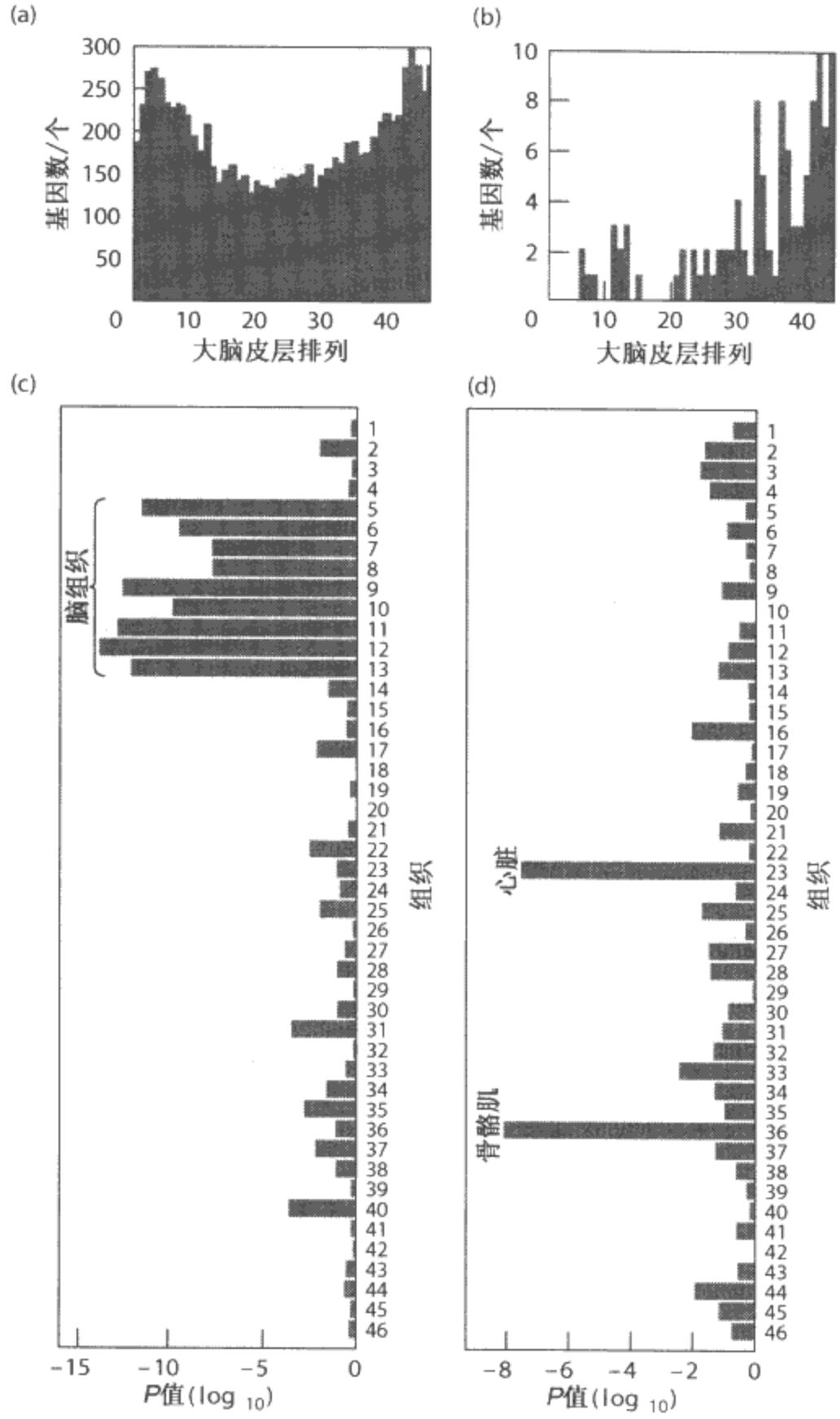


图 24.23　RNAi 所致组织特异性基因表达下调。（a）大脑皮层中基因表达的排列。46 种组织 10 000 个人类基因在图中排列如下：最左边的柱代表在大脑皮层比在任何其他组织表达都高的基因，下一个（左边第二个）柱代表除上一个之外在大脑皮层比在任何其他组织表达都高的基因，而最后一个（最右边的）柱代表在大脑皮层比在任何其他组织表达都低的基因。（b）miR-124处理后 mRNA 水平明显减少的基因。请注意，以（a）图为背景的显著性 P 值约为 10^{-12}，柱形图明显地偏向在大脑皮层中低表达的基因。（c）将（b）图的分析（miR-124 影响基因表达作用的分析）扩大到 46 种组织，并对 P 值的对数作图。只有脑组织的 P 值（<0.001）呈显著性差异。（d）类似于（c），只是分析的是经 miR-1（而不是 miR-124）处理的细胞。（*Source*：Adapted from Lim et al.，Microarray analysis shows that some micro RNAs downregulate large numbers of target mRNAs. *Nature*. Vol. 433（2005）f. 1，P. 770.）

小结　通过检测被外源性miRNA消减的mRNA谱并将其与各种组织中在mRNA水平上的基因表达谱比较，可以开展组织特异性表达分析的研究。假如一 miRNA 细胞中 mRNA 的表达降低，则证明这些 miRNA 在生理条件下亦能在一定程度上导致这些 mRNA 的低表达，这类分析暗示 miR-124 破坏了脑组织中一些 mRNA 的稳定性，而 miR-1 破坏了肌肉组织中一些 mRNA 的稳定性。

转录因子靶位点的定位 正如我们在第 12 章中所了解到的，基因被结合在增强子上的激活因子所激活。在一个基因组中很多激活因子具有多个增强子标靶，因此能激活很多基因。有时将一组可同时受到调节的基因称为**调节子**（regulon）。为了完全搞清给定激活因子的效应，有必要对某个激活因子有响应的所有基因进行鉴定，目前，通过几种方法可以开展此项工作。

最为直接的方法是比较一激活因子导致的不表达、低表达和高表达生物的 RNA 微阵列的杂交模式。从这一分析中我们可获知由于激活因子高表达所激活的基因，此方法对于这一目的实现有着重要的作用。但还是有两个问题限制其应用。其一，被激活的基因也许并不是第一个激活因子直接的标靶，而是第一个激活因子激活的另一个激活因子的标靶；其二，被激活因子过量表达所激活的基因不一定能够被体内生理浓度的激活因子激活。尽管如此，还是有一些方法通过直接检测激活因子与特定基因调控区的互作来解决这些问题。

其中一种策略是 Richard Young 及同事（Ren et al.，2000）使用过的，融合了染色质免疫沉淀（ChIP，第 13 章）和 DNA 微阵列杂交分析两种不同技术。图 24.24 说明了该方法的总体方案，这是 Young 及同事在酵母全基因组范围内鉴定激活因子 GAL4 的结合位点时采用的方法。首先，利用化学方法将蛋白质交联到染色质 DNA 上以避免二者分离；然后破碎细胞并把染色质剪切成小片段；接着用抗 GAL4 抗体免疫沉淀剪碎了的酵母染色质以沉淀结合有 GAL4 的 DNA；之后将交联的 DNA 与蛋白质解离，并通过 PCR 以红色荧光染料（Cy5）标记这一 DNA。同样的方法以绿色荧光染料（Cy3）标记未被免疫沉淀的 DNA。再用这两种标记的 DNA 做探针，与含酵母全基因组基因间区段的 DNA 微阵列杂交。图 24.25 显示的是此阵列的一小部分。其中一个由箭头标示出的点清楚地显示出红色荧光信号占优势，说明此点优先与 GAL4 结合的 DNA 杂交。通过这一技术，Young 及同事鉴定了 10 个基因序列，这些都是被 GAL4 所激活的基因。因此，该方法在这一实验中非常有效。

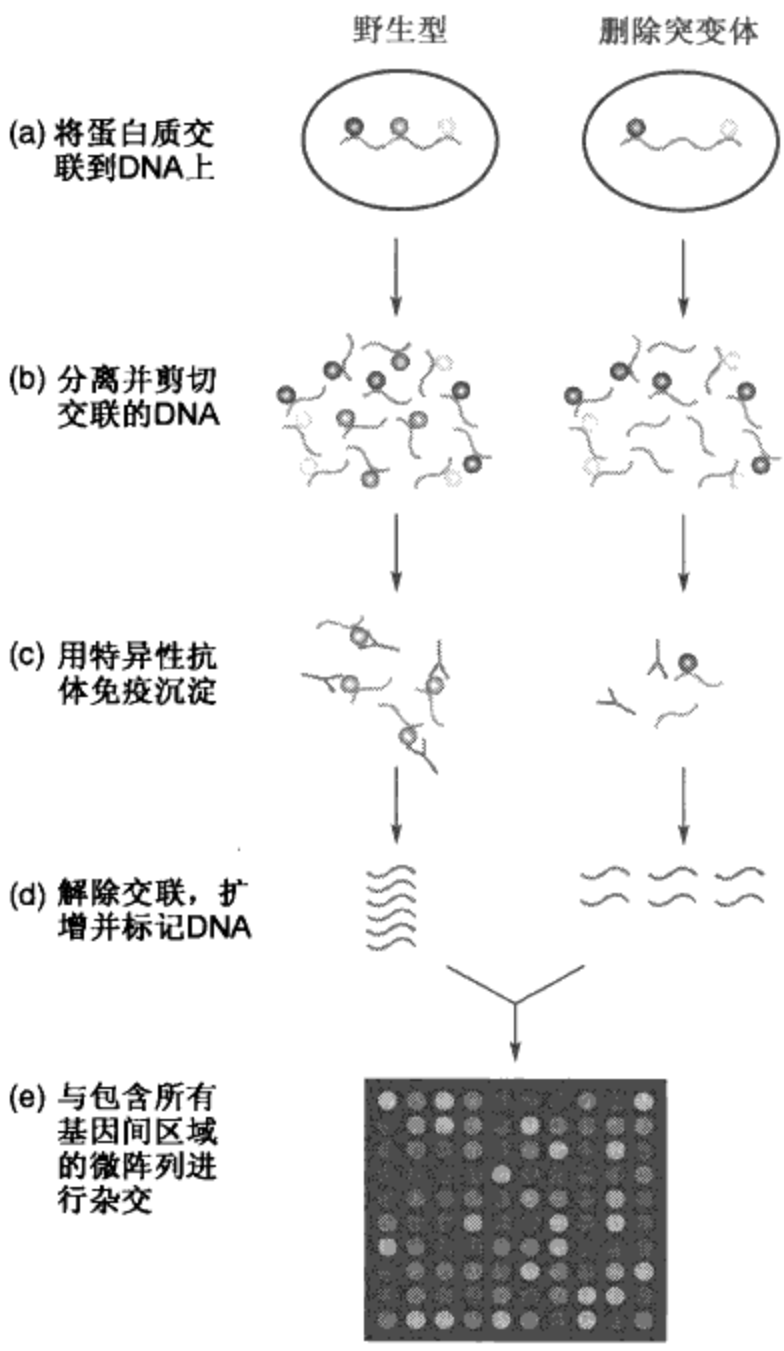


图 24.24 在酵母的全基因组中寻找 DNA-蛋白质间的互作。（a）首先，利用化学方法将蛋白质交联到酵母细胞 DNA 上，酵母细胞包括野生型和缺失了靶蛋白编码基因的参考细胞株（红色）。（b）通过破碎细胞，将蛋白质-DNA 复合物（与染色质交联）从细胞中分离出来，并进行超声波剪切。（c）用靶蛋白的抗体免疫沉淀剪切的 DNA。（d）沉淀后，解除交联，PCR 扩增并标记沉淀的 DNA。（e）两种类型细胞中标记的 DNA 与包含酵母基因组中所有基因间 DNA 的微阵列进行杂交，野生型细胞中沉淀的 DNA 用红色的荧光染料标记，突变型细胞中沉淀的 DNA 用绿色的荧光染料标记，因此，如果微阵列上的一个 DNA 位点与靶蛋白结合的 DNA 杂交强于与其他蛋白质结合的 DNA，此位点呈现红色荧光。如果优先与其他蛋白质结合的 DNA 杂交，位点呈现绿色荧光。如果与上述两种 DNA 都结合，将呈现黄色荧光。对两种 DNA 探针相对荧光强度进行仔细校准，使我们能够确定每一个位点红色和绿色荧光的比率，因而靶蛋白结合的 DNA 区域被优先选择出来。（*Source*：Adapted from *Nature* 409：from Lyer et al.，2001，Fig. 1，p. 534.）

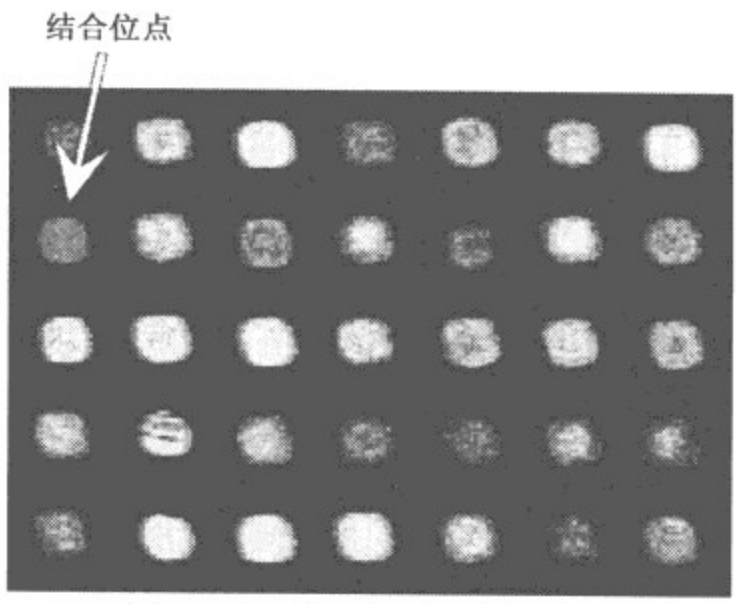


图 24.25　鉴定与 GAL4 结合的一段 DNA 序列。Young 及同事用 PCR 扩增抗 GAL4 抗体免疫沉淀的染色质上的 DNA，制备一个红色荧光的 DNA 探针。同时用 PCR 扩增并非抗体免疫沉淀获得的 DNA，制备一个类似的绿色荧光的 DNA 探针。然后将两种探针与包含酵母基因组所有基因间 DNA 的微阵列杂交。该图是阵列的一小部分，红色的位点（箭头表示）显示一个推定的与 GAL4 结合的 DNA，绿色位点显示与 GAL4 不结合的 DNA，几个黄色的位点（与红色和绿色的探针都结合）显示与 GAL4 结合并没有显著优先的 DNA。（*Source*: Adapted from Ren et al., *Science* 290 (2000) Fig. 1A p. 2306.）

这一方法之所以在酵母中非常有效是因为酵母的基因组较小且已完成全序列测定。但类似的方法能应用于人类基因组吗？那就存在很多问题了，因为人类基因组中基因间的部分几乎与基因组本身大小相同，含人类基因组所有基因间序列的微阵列分析非常复杂也很难实施。然而也有一些方法可以缩小 DNA 序列分析的范围以使实验便于操作，其中的两个均为 2002 年报道的有关人类 E2F4 激活因子。

Peggy Farnham 及同事缩小分析范围的方法是进行仅含 CpG 岛（7776 个）的微阵列分析。正如我们在本章前面所了解的，CpG 岛与基因调控区相关，因此也富含该技术寻找的激活因子结合序列。利用这一策略，Farnham 及同事鉴定出激活因子的 68 个标靶位点。David Dynlacht 及同事没有选择 CpG 岛，而是选择已知在细胞进入细胞周期（即 E2F4 活跃期）时被激活的约 1200 个基因的调控区作为研究对象。通过这一组 DNA 的微阵列分析，他们在人纤维原细胞中发现了 127 个与 E2F4 结合的序列。因此，提前了解一些激活因子定时或选择性表达方面的知识，有助于设计筛选靶基因的微阵列。

小结　染色质免疫沉淀后的 DNA 微阵列分析能够鉴定激活因子和其他蛋白质的 DNA 结合位点。对基因组较小的生物体（如酵母）而言，微阵列分析包括了所有基因间的区域。但对大基因组而言（如人类基因组）这无法实施。CpG 岛可用于缩小微阵列分析的范围，因为其与基因的调控区域相关。如果一个激活因子的活性作用时间和条件已经知道，则在这些时间或条件下处于激活状态的基因的调控区也可用于此目的。

原位表达分析（*in situ* expression analysis）考虑到已经知道人类 21 号染色体与唐氏综合征有关，我们就有机会在该染色体中发现导致此疾病的基因，这将有助于了解这个染色体上所有基因在胚胎期的表达模式。

对于较低等的生物体而言，这类研究可按常规实施，经典的是用 cDNA 探针对胚胎切片进行原位杂交（第 5 章）。但存在一个严重的问题，即如果使用人的胚胎开展这类研究将涉及伦理上的问题。所幸的是我们已经获得了小鼠基因组全序列，因而有办法绕过这一问题。人类 21 号染色体上已确定的基因有 178 个，其中 161 个在小鼠基因组中存在直系同源物（ortholog）。所以我们可以全程、全方位追踪小鼠胚胎发育过程中这些基因的表达，而且我们设想人类胚胎中的同源基因采用相似的基因表达模式。

两个研究小组将这一策略应用于人类 21 号染色体所含基因在小鼠中的同源基因上。其中之一是 Gregor Eichele、Stylianos Antonarakis、Andrea Ballabio 及同事，通过原位杂交法对 158 个小鼠同源基因在妊娠期间 3 个不同时期的表达情况进行了调查。他们还利用 RT-PCR（第 5 章）检测了所有 161 个同源基因在成体中的表达模式。结果发现了几个基因局限于特定时间和特定位点的表达模式，而且，其中一些模式表达的部位（中枢神经系统、心脏、胃肠道、肢体）与唐氏综合征的病理学相一致。

例如，图 24.26 显示了 *Pcp 4* 基因在小鼠胚胎第 10.5 天（全部封固切片的原位杂交）和第 14.5 天（胚胎切片的原位杂交）时的表达。在 10.5 天，基因在眼睛（黑色箭头）、大脑和背根神经中枢（白色箭头）中表达。在第 14.5 天，基因在许多组织中均表达，包括大脑皮层组织（红色箭头）、中脑、小脑、脊髓、肠、心脏和背根神经中枢，所有这些部位都是唐氏综合征所影响的区域。所以，*Pcp 4* 基因是与此疾病有关的候选基因之一。

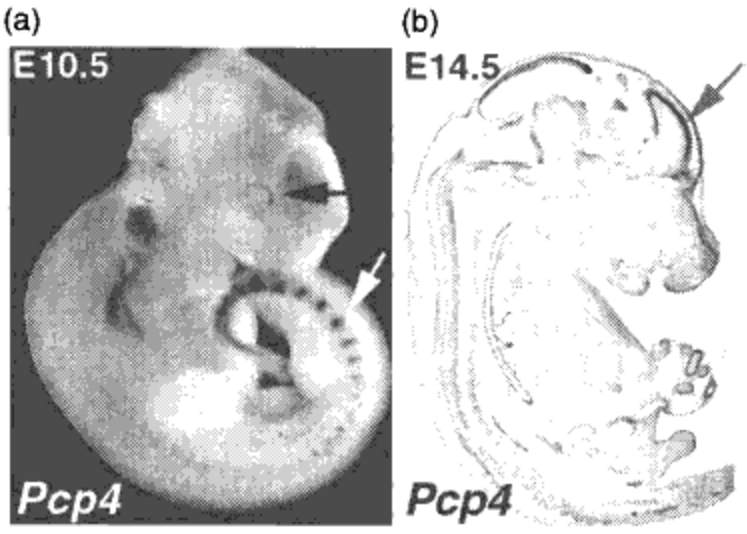


图 24.26　两个基因在小鼠胚胎中的表达。使用全封固胚胎（a）或切片胚胎（b）的原位杂交（第 5 章）检测基因的表达。（a）*Pcp 4* 在 10.5 天的封固全胚胎中的表达。黑色箭头指出了眼睛，白色箭头指出了背根神经中枢。（b）Pcp4 基因在 14.5 天的胚胎切片中的表达，红色箭头指示大脑皮层板，深色区显示基因的表达。（*Source*：Adapted from *Nature* 420：from Reymond et al.，fig. 2，p. 583，2002.）

另一个例子是整合了 Eichele 及同事，以及由 Ariel Ruizi Altaba、Bernhard Herrmann 和 Marie-Laure Yaspo 率领的研究组的工作，是关于小鼠 *SH3BGR* 基因在妊娠第 9.5 天、第 10.5 天和第 14.5 天的表达。研究结果显示 3 个发育阶段此基因在心脏中都有显著的表达，因为心脏是唐氏综合征影响的器官之一，所以，*SH3BGR* 基因是与此疾病有关的另一个候选基因。

小结　小鼠可以作为人类大规模基因表达研究的替代品，因为就伦理而言，不可能用人做这方面的研究。例如，科学家已经研究了小鼠几乎所有的和人 21 号染色体上同源的基因的表达，并且追踪了这些基因在胚胎发育不同阶段的表达，列出了表达这些基因的胚胎组织名录。

单核苷酸多态性：药物基因组学

既然我们已经获得了人类基因组序列的草图，就能够寻找个体间的差异了。迄今为止，大部分的差异存在于单核苷酸，我们称它为**单核苷酸多态性**（single-nucleotide polymorphism，SNP）（**SNP** 发音为“snips”）。如果能将这些 SNP 与单个基因缺陷而造成的人类疾病联系起来，就能够发现那些有可能患这些疾病的个体，也可能会发现一套多基因性状的 SNP，如对心血管疾病和癌症敏感的 SNP，从而阻止这些基因的作用。

我们还可以鉴定一些对某种药物起反应或不起反应的相关的 SNP。利用这些信息，医生可以筛选出患者中的关键 SNP，并根据患者 SNP 对药物的反应，为患者制定一个药物治疗程序，这一领域的研究称为**药物基因组学**（pharmacogenomics）。

但是，要完成这些目标是不容易的，已有的研究表明，大部分 SNP 根本不存在于基因内，而位于基因间隔区内，它们中的大部分都不会影响基因的功能。但是，如果被定位在基因控制区域内，则会影响基因的功能。甚至，即使发现它们存在于基因内部，有时也倾向于沉默突变，不会改变蛋白质产物的结构，因此，不会引起任何导致疾病的功能障碍。出现这种情况的原因是：改变基因产物的突变引起的多态性通常是有害的，因此不会被选择。那些具有危害性的突变个体，在它们能够繁殖前，通常已经死亡，所以那些突变也就丢失了。

我们可以通过各种基因分型技术，发现与一个给定个体的疾病或其他特性相关的 SNP。一种方法是将一引物杂交到 SNP 相邻的一段序列上，然后，用荧光标记的核苷酸进行引物延伸，观察哪个核苷酸将掺入 SNP 位点。另一种方法是将某个人的 DNA 与一个包含野生型和突变型寡核苷酸序列的 DNA 微阵列杂交。还有一种方法是通过 PCR 扩增一段围绕 SNP 的序列，然后进行测序。这些知识将有益于人们预防和治疗疾病。但是，也同知识窗 24.1 中指出的，它们也是有弊端的。

SNP 和 RFLP 之间有什么不同？如果两个体之间单核苷酸的不同在于一个限制酶位点上，那么，它们两者是一致的，就像我们在本章前面所提到的，在亨廷顿病患者中，RFLP 与 *Hind* III 酶切位点有关。此例中，一个单核苷酸的不同，使得限制性片段的模式发生了改变。然而，RFLP 也可以由一个个体的两个限制性位点之间插入一段 DNA 而形成，但不是 VNTR，因为它所涉及的不仅仅是单核苷酸的不同。

由于人们热心于 SNP 在鉴定普通疾病原因方面的潜能，2005 年，国际单倍体型图谱联盟（International HapMap Consortium）公布了一张包括一百多万个人类 SNP 的单倍体型图谱，通过对来源于 4 个截然不同的人群（一个在尼日利亚、一个在美国犹他州、一个在中国、一个在日本）的 269 个 DNA 样品进行基因分型后发现的。**单倍体型图谱**（haplotype map）显示了单倍体型的位置及倾向于完整遗传的 DNA 片段（其内部重组率很低）。在有关人类基因组的讨论中，我们已经了解到不同位点的重组率变化相当大，且高重组率的区域与低重组率的区域交替出现。低重组率的区域很可能包含一些一起遗传并组成单倍体型的遗传标记。

通过集中一些选定的（well-chosen）SNP［标签SNP（tag SNP)］，国际单倍体型图谱联盟能够在同一区域鉴定出另一些SNP，因此，就基因分型的总数而言，减少了工作量。通过将标记的人类DNA片段与专门检测标签SNP的DNA微阵列杂交来进行大规模的基因分型。此方法已高度自动化，一个人仅用两天时间即可完成对全基因组上500 000个SNP的筛选。

这一计划的直接成果是鉴定了几百万个新的SNP（在该计划开始之前我们知道的SNP只有1 700 000个）。另一个直接的成果是我们对人类进化过程中的重组和自然选择有了新的认识。

而潜在的（也是最吸引人的）成果是鉴定人类疾病相关基因。这种鉴定在HD和其他由单基因突变引起的疾病中很直接，因为有特定突变的人几乎一定会患相应的疾病。但是如果多基因在疾病发生中起作用就很难了，因为每个基因的突变只发挥一点作用，因此每个基因的发现都很困难。不幸的是，许多人的致死或致残疾病（如癌症、心脏病和痴呆）属于后者。原则上，HapMap能使这项工作容易些，但正如前面已经提到的，疾病相关基因的鉴定并不能立即导致有效的治疗。在不久的将来，HapMap最可能用作药物基因组学的工具。

另一些研究者的“视野”已经超出SNP，而是比较不同人的基因组，并且得到了一些令人吃惊的结果。表面上为正常人的基因组中，常常不仅含有SNP，而且也含有缺失、插入、倒位和完整的大段DNA的重排。现在，遗传学家渐渐开始称基因组中的这类差异为**结构变异**（structural variation）。例如，Michael Wigler及同事检测了20个健康个体的基因组，并在221处发现特定DNA片段有不同的拷贝数，而这些拷贝数的变化并没有明显地影响他们的健康。也许结合某些环境因素，这些变异可能会使其他人患病。

此外，一些结构变异似乎是有益的。Sunil Ahuja及同事证明特定的一些免疫系统基因的额外拷贝有助于人抵抗AIDS。冰岛的一个科学家团队发现20%的欧洲人携带一个大的倒位。令人惊讶的是，携带这一倒位的妇女比不携带的这一倒位的妇女有更多的孩子，暗示这一倒位赋予了某种生殖方面的优势，所以可能会更广泛地传播。

简单生物体的全部基因组序列对于理解和治愈人类疾病也很重要。例如，酵母基因组的序列完全测定后，分子生物学家就开始系统地突变6000个酵母基因，调查这些突变会导致什么样的结果。通过酵母双杂交筛选（第5章及本章的后面），他们也开始系统地筛查可能的1.8×10^7个蛋白质—蛋白质间的互作效应。这类实验的结果能使我们对未知基因产物的活性有更多的了解。知道了一个生物体中的所有蛋白质的活性以及与它们互作的其他蛋白质的性质，使我们能更加深入地理解一些生化途径，如药物的代谢途径、控制基因表达的信号传导途径。这种了解又为我们对人类的类似途径的理解提供了重要线索。此外，酵母细胞可以用作人类的替代品来检测敲除酵母中的和人类已知疾病相关的同源基因所造成的影响。

小结 单核苷酸多态性能够解释许多由单基因甚至多基因控制的遗传疾病。它们也许能够预测人对药物的反应。含一百多万个SNP的单倍体型图谱使我们很容易从没有影响的SNP中挑选出重要的SNP。结构变异（缺失、插入、倒位和完整的大段DNA的重排）也是人类基因组变异的很重要来源。理论上，一些结构变异可能会使某些人更易患病，有些则是与健康关系不大，而有些则显示是有益的。

24.4 蛋白质组学

在本章的开始，我们了解到对某个生物一生中所有蛋白质性质和活性的研究，称之为**蛋白质组学**（proteomics）。分析生物的基因组，甚至转录基因组的工作是相对直接的，然而分析其蛋白质组则不容易，很大程度上是因为蛋白质比核酸复杂。靠目前的技术，复杂生物体蛋白质组学仅能检测蛋白质组的一部分。

既然难度如此之大，为什么科学家在已获得转录组学方面的信息，而且可以根据转录的水平同时检测到大量基因表达的情况下，仍然对蛋白质组学如此感兴趣呢？部分原因是我们现在知道在人类细胞中很大比例（50%或者更多）的有多聚腺苷酸尾的RNA并不编码蛋白质，被称为非编码RNA（noncoding RNA，**ncRNA**）。而且，正如我们已经看到的，它们也被称为未知功能转录物（TUF）。虽然TUF是令人感兴趣的，但它们的表达水平并不能告诉我们蛋白质的表达水平。另一个原因是蛋白质编码基因的序列及其表达水平并不能说明蛋白质产物的活性问题。

还有，基因的转录水平只能给出该基因表达水平的大致情况。一方面某种mRNA可能大量合成，但可能立刻被降解或低效翻译，结果蛋白质产量很少。另一方面有很多蛋白质要经历转录后的修饰，这对它们的活性有着深刻的影响。例如，有些蛋白质直到磷酸化后才有活性。所以如果细胞没有将这个蛋白质在适宜的时候磷酸化，再多的mRNA也不

会显示出这个蛋白质的真正表达水平。不仅如此，很多的转录产物会通过选择性剪接及有选择地转录后修饰产生不止一种蛋白质。因此，仅看基因的转录并不一定能够知道它会产生什么蛋白质。最后，很多来自大的复合物的多肽的真正功能是以整体的活性为基础的。

所以，如果要检测真实的基因表达，必须看蛋白质的水平。要分析某生物体的所有蛋白质，必须做两件事：第一，把所有的蛋白质分离出来；第二，鉴定所有蛋白质并检测其活性。下面的两节我们将了解分子生物学家是怎样做这两件事的。

小结 一个生物的所有蛋白质产物的总和称作蛋白质组，研究这些蛋白质或其中很小的一部分即为蛋白质组学。这类研究对基因表达的描述比转录组学研究更为精确。

蛋白质分离

最好的分离工具是发明于 20 世纪 70 年代的双向电泳（two-dimensional gel electrophoresis，**2-D**）（第 15 章）。这种技术的功能还不能达到分离人类所有蛋白质组的水平。普通的 2-D 胶仅能分离大约 2000 种蛋白质，即使用最好的胶和最好的技术也只能分离大约 11 000 种蛋白质。事实上，双向电泳的结果是不可预测的，仿佛是一种艺术而不是科学，这又增加了其难度。还有一个问题是，很多有趣的膜蛋白由于太疏水而不能溶于电泳缓冲液中，因此在电泳中是看不到的。最后，很多蛋白质在细胞中的量很少，以致双向电泳检测不到它们的存在。现存的很多问题是相当棘手的，但是科学家通过分步处理的方法来解决这些问题。例如，他们先从细胞核甚至是其亚结构核仁或者核孔复合体蛋白的分离着手。随着更多的蛋白质被分离，问题就没有那么困难了。

蛋白质分析

一旦蛋白质被分离并定量了，如何分析呢？首先，需要进行鉴定。这项工作现在最好的方法是：将分离出的各个斑点切出，用蛋白水解酶裂解成肽段，这些肽段可用**质谱分析**（mass spectrometry）鉴定。图 24.27 展示了一个目前很通用的但却有一个复杂名称的技术——基质辅助的激光解吸电离飞行时间质谱分析系统。MALDI-TOF（matrix-assisted laser desorption-ionization time-of-flight）。在这个程序中，肽段被固定于基质上形成晶体。接着，肽段被激光束离子化（基质帮助肽段离子化），基质电压的升高导致离子射向一个检测器。假设所有离子的带电情况相同（通常几乎如此），那么离子到达检测器的时间就取决于离子的质量。质量越大，离子飞行的时间越长。在 MALDI-TOF 质谱分析仪中，离子被能产生离子束的静电反应器偏转。所以，我们可以精确地估计到达第二个检测器的离子质量，这些质量可以揭示肽段的准确构成以及属性。如果整个基因组序列已知，我们知道蛋白质的属性，计算机就可以运用来自质谱分析的信息将双向电泳的结果和基因组的信息对应，预测蛋白质的全序列。当然，知道蛋白质序列并不一定知道其功能，因此有必要进一步研究确定蛋白质的功能。

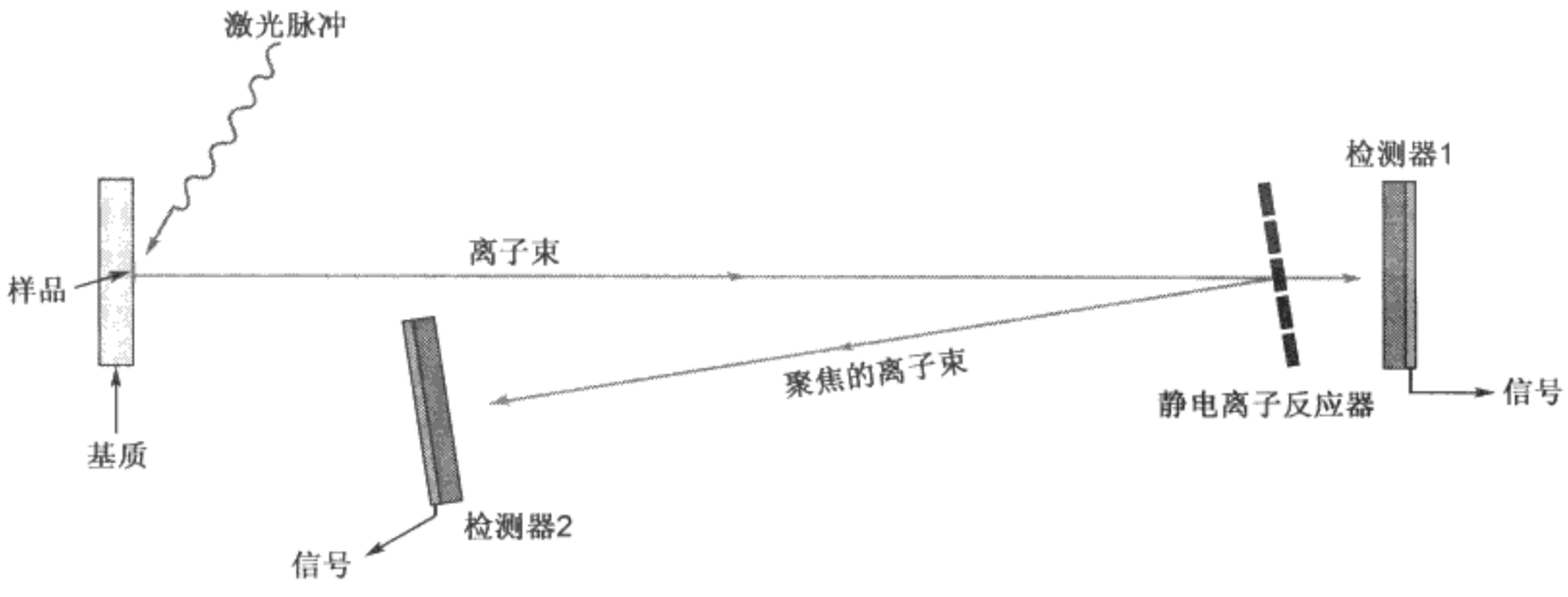


图 24.27 MALDI-TOF 质谱原理。将样品（本例为肽段）置于基质上，加一个激光脉冲使其离子化。基质和样品间的电势差异使离子化的样品向检测器 1 加速。离子到达检测器 1 的时间取决于离子质量，可以根据飞行时间来分析样品分子的质量。还可以选择性地打开检测器 1 前面的静电离子反应器，以集中离子并将其反射到检测器 2。这个检测器能够根据飞行时间计算出更精确的离子质量。

你可能会想：用微芯片来鉴定成千上万的蛋白质是不是会更好——就像功能基因组学研究中用 DNA 芯片一次检测数以千计的 RNA 一样。那样就无须分离蛋白质了，因为很多蛋白质的混合物与芯片孵育后

就能观测到是何种蛋白质结合了。这关系到一个策略：生产能够特异并定量地识别蛋白质的抗体，并将其固定于芯片上。但是要实现这一梦想还有许多障碍。首先，抗体非常昂贵，制备起来也比寡聚核苷酸耗时得多。事实上，制备人的每一个蛋白质的抗体的工作量之大是无法想象的。而且，探测低含量的蛋白质对双向电泳来说已不可能，小型的芯片技术更是加剧了这项工作的难度。然而，80 年代中期人们最初提出花一段时间用完成人类基因组计划，当时在技术上是不可能的，但是此计划促进了该技术的发展。或许，如果我们发起人类全蛋白质组计划，可能会经历与全基因组计划相同的历史。

小结 当前的蛋白质组学研究首先要求分离蛋白质，有时是大规模的分离。最好的对大量蛋白质进行同时分离的工具是双向电泳（2-D）。蛋白分离之后必须鉴定，最好的方法是用蛋白酶切成肽段，再用质谱分析鉴定这些肽段。有朝一日，带有抗体的蛋白质芯片也许会取代分离法而分析复杂的蛋白质混合物。

蛋白质的相互作用

很多蛋白质单独没有功能，但是会和其他蛋白质相互作用。许多蛋白质相互作用一起参与生化或发育途径。信号传导途径是最好的例子。多种蛋白质一起形成多蛋白质复合体执行一个特殊的任务，如核糖体（蛋白质合成）或者蛋白酶体（蛋白质降解）。所以蛋白质组学的任务是鉴定相互作用的蛋白质。这常常为推测新蛋白质的功能提供重要的线索。

探究蛋白质之间相互作用的传统方法是应用酵母双杂交分析（第 5 章）（译者注：第 14 章），一些蛋白质组水平的蛋白质-蛋白质相互作用的研究也是由此技术演绎而来的。但是双杂交分析是间接的，它是应用报告基因的激活来观察嵌合的转录激活因子的两部分之间的互作效应，而且受假阳性和假阴性的影响。然而，如能与一个依赖其他技术的验证实验相结合，酵母双杂交筛选体系还是非常高效的。2005 年，Frich Wanker 及同事用酵母双杂交体系（部分结合了独立证实）检测到 3000 多种人类蛋白质间的相互作用——减缓这一艰辛路途并成为向阐明**互作组**（interactome）迈进的起点，互作组是指所有人类蛋白质间的相互作用。

近来有些研究者开始用超敏感的蛋白质谱分析，更有效地检测蛋白质-蛋白质间的互作。

有关此方面研究的实例是 2002 年 Daniel Figeys 及同事采用以下程序（图 24.28）来检测酵母中蛋白质-蛋白质的互作：首先，他们筛选出一套（725 种）“诱饵蛋白”（bait protein），它们很容易与“鱼

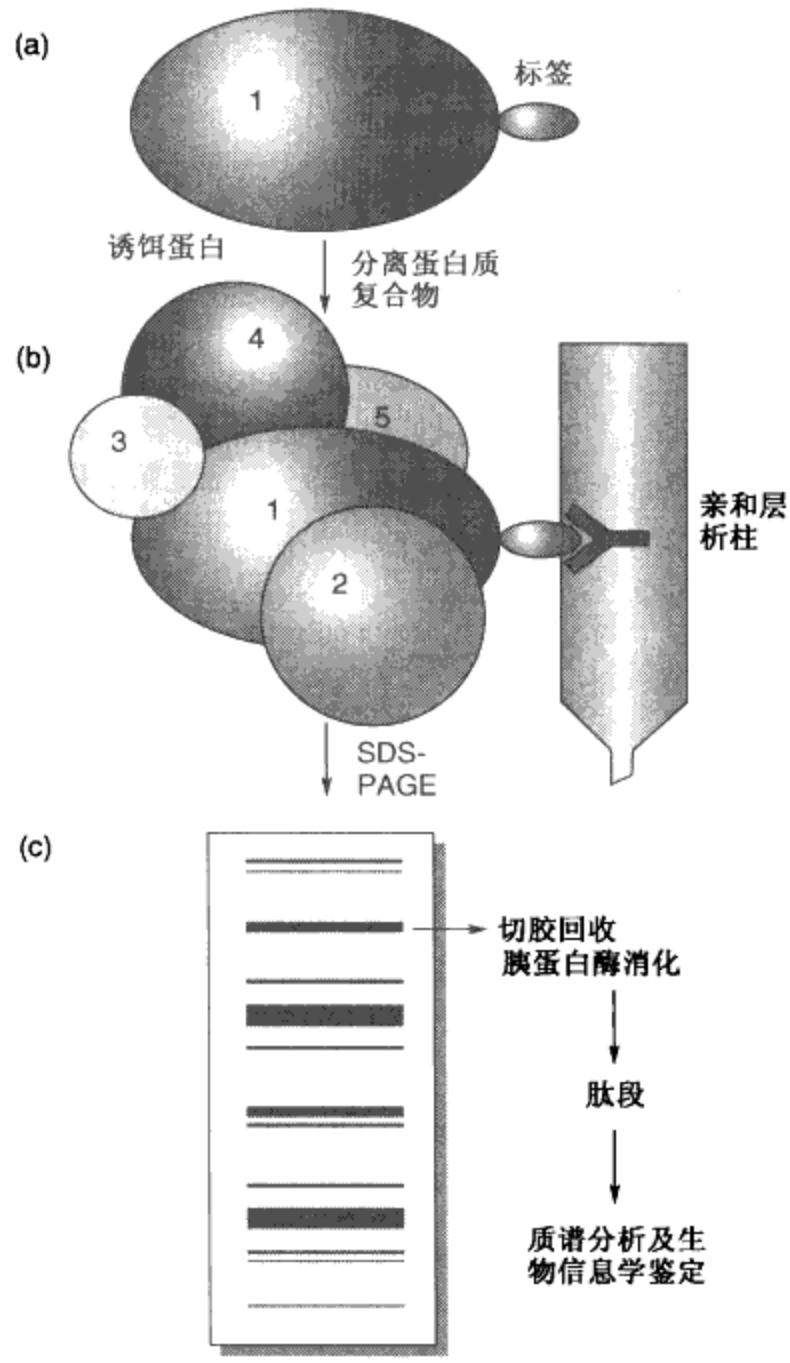


图 24.28 用质谱检测蛋白质-蛋白质互作。（a）制备带标签的诱饵蛋白。一个编码诱饵蛋白的酵母基因经基因工程改造后加入标签（如 Flag 表位）编码序列，然后导入酵母细胞表达带标签的诱饵蛋白。（b）应用免疫亲和层析（所用的树脂带有抗标签抗体）分离带有诱饵蛋白的复合物。这样不仅能“钓出”诱饵蛋白，而且能“钓出”所有与诱饵蛋白作用的蛋白质。此实例中有 4 种蛋白质：2～5 号。（c）纯化并鉴定蛋白质。应用 SDS-PAGE 分离纯化复合物中的蛋白质。切取凝胶上的蛋白质，用胰蛋白酶消化，生成的肽段再用质谱分析。计算机将胰蛋白酶消化生成的肽段质量与根据所有酵母基因组编码蛋白预测的多肽质量进行比较，以鉴定蛋白质。（*Source*：Adapted from Kumar, A. and M. Snyder, Protein complexes take the bait. *Nature* 415, 2002, p. 123, f. 1.）

蛋白”（fish protein）作用。这些“诱饵蛋白”代表不同的类群，包括蛋白激酶、磷酸化酶和一些参与到损伤 DNA 修复的酶。研究者把以上各类蛋白中的每一种所对应的基因进行了克隆（包括 Flag 抗原

表位基因），然后将这个嵌合基因导入酵母细胞并使其表达（“Flag”一词是指该抗原表位作为“旗帜”使蛋白质易于被单个抗体识别）。

然后，研究者用带有抗 Flag 抗体的免疫亲和层析柱从细胞提取物中分离纯化出含有诱饵蛋白的蛋白质复合体。他们用 SDS-PAGE 将复合物中各个蛋白质分开，切下凝胶的各条带，裂解每条带的蛋白质，然后分析裂解后的肽段。因为我们知道酵母的整个基因组，所以计算机可以预测出基因组编码的任何蛋白质和裂解后肽段的质量。所以，这种生物信息学方法（见下一节）可以利用质谱分析仪的数据来鉴定胰蛋白酶裂解后的肽段，进而分析蛋白质。

用 10% 的已预测的酵母蛋白质作“诱饵”，Figeys 及同事“钓”出并鉴定了 3617 种相关蛋白，大约包括酵母蛋白质组的 25%。这是酵母双杂交体系的 3 倍还多。图 24.29 显示用两种蛋白激酶 Kss1 和 Cdc26 作诱饵的实验结果。有些已知互作（红色箭头）被重新发现，很多未知互作（绿色箭头）也首次被找到了。

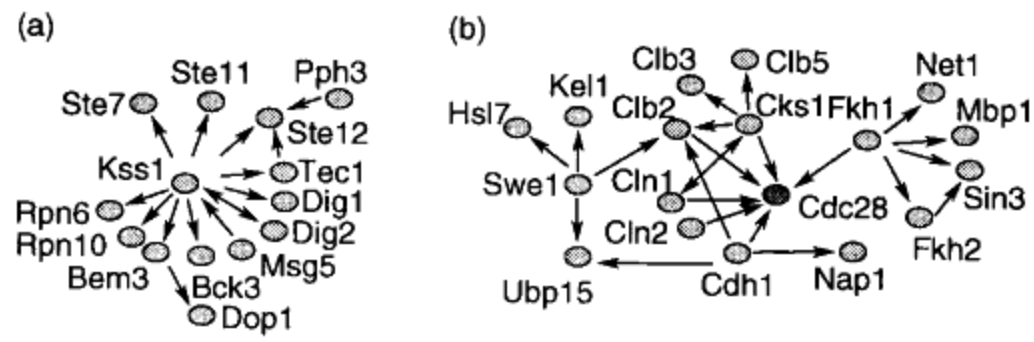


图 24.29　Figeys 及同事发现的蛋白质-蛋白质互作的实例。（a）以 Kss1 为诱饵发现的蛋白质互作。（b）以 Cdc28 为诱饵发现的蛋白质互作。红色箭头指示已知的蛋白质互作，绿色指示该研究中发现的蛋白质互作。（*Source*: Adapted from Ho, Y., A. Grahler, A. Heilbut, G. D. Bader, L. Moore, S. L. Adams et al., Systematic identification of protein complexes in *Saccharomyces carevisiae* by mass spectrometry. *Nature* 415, 2002, p. 180, f. 1.）

在类似研究中，Anne-Claude Gavin 及同事在酵母 232 个多蛋白复合体中发现了 589 种蛋白质组装。更有趣的是，这些结合可以预测 344 种蛋白质的新功能，其中有 231 种的功能在以前是未知的。这种“连带”（guilt by association）技术对发现未知蛋白功能是一个有力的工具。

Michael Snyder 及同事从另一角度解决这个问题。他们用载有酵母蛋白质组的微阵列来检验哪些蛋白质（或脂质）能与微阵列的哪个蛋白质结合。微阵列上的每个小点含有一个酵母蛋白，其上耦合了 GST 和一个寡聚组氨酸标签。事实上，蛋白质通过组氨酸标签与镍包裹的芯片结合。在用此法的一个实验中（图 24.30），Snyder 及同事用一个带有生物素的蛋白质或脂质作探针检测芯片，再用带荧光尾巴的链霉亲和素检测。链霉亲和素（streptavidin）与生物素紧密结合，其尾部发绿色荧光表明此处显阳性。微阵列中的蛋白质点阵具有一个重复，所以阳性应显现成对的绿点。图 24.30 显示的每一张片子至少有一个阳性反应。钙黏蛋白是一种钙结合蛋白，在钙离子存在的条件下能与其他很多蛋白质反应。其他 5 种探针是带有生物素的脂质体，它们大部分都有细胞内信号传递活性。微阵列同时也由抗 GST 抗体和发红色荧光的二抗标记。这可作为一个对照实验；所有的蛋白质都有一个 GST 标签，所以它们都应该能被 α-GST 抗体“点亮”。

有些蛋白质对其他蛋白质的某些特殊序列有结合模块。例如，SH3 域和 WW 域结合富脯氨酸肽段，SH2 域结合含磷酸酪胺的肽段。基于这一点，Stanley Fields、Charles Boone、Gianni Cesareni 及同事建立了一种实验方案——结合实验和计算机手段鉴定能与这些蛋白质互作的元件，当然这些蛋白质有上述或其他的肽结合区域。

该方案分以下 4 个步骤：第一，研究者应用**噬菌体展示**（phage display）技术发现了与一定的肽结合域识别的共同序列。在噬菌体展示技术中，编码蛋白质或肽段的基因或基因片段被克隆到噬菌体载体上与噬菌体衣壳蛋白基因融合以便该蛋白或肽段展示于重组噬菌体的表面。

在这个研究中，Tong 及同事以原癌蛋白 Src 作查询序列，用 Ψ-BLAST 分析（见下一节）鉴定出酵母中 24 种不同的 SH3 域。其中有 20 种 SH3 可以在 *E. coli* 中表达为 GST-融合蛋白。Tong 及同事从一个展示在噬菌体表面的随机九肽（random nonapeptide）文库筛到了 SH3。每个 SH3 都优先结合一组九肽，由此鉴定出每个 SH3 结构域的靶肽的共同序列。

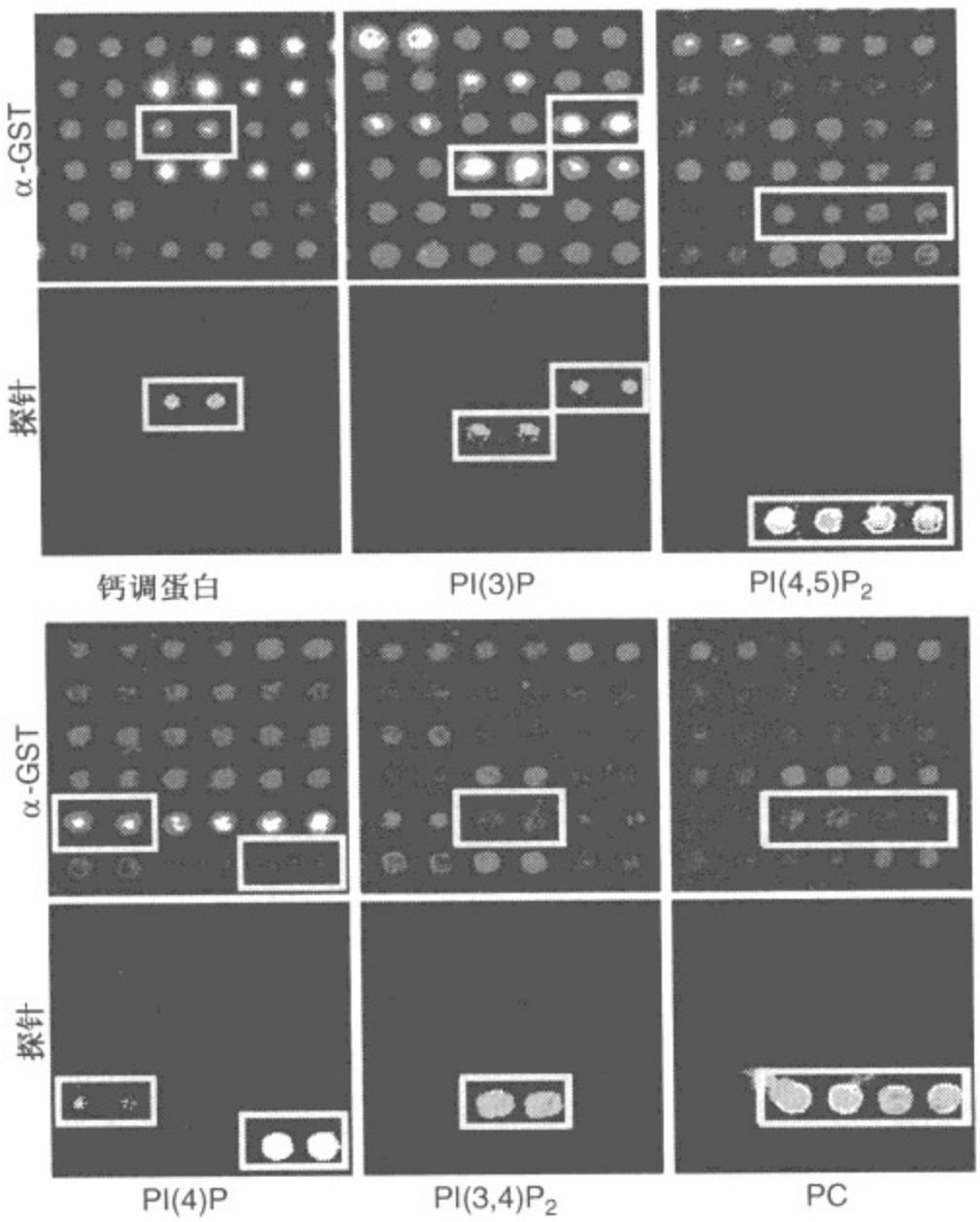


图 24.30　用蛋白质芯片检测蛋白质-蛋白质的互作和蛋白质-脂质的互作。Snyder 及同事用并排重复方式构建了蛋白质微阵列，首先用 α-GST 抗体（第一排和第三排）或不同探针（下方的第二排和第四排）进行检测。然后再用荧光探针检测 α-GST 抗体从而产生红点。红色荧光的强度反映了每一点中蛋白质的量。第二排和第四排的探针偶联了生物素，可用偶联了绿色荧光标签的链霉亲和素对其进行检测。探针是钙调蛋白（钙调蛋白是一种在很多代谢途径中起作用的蛋白质，它需要钙）和带有以下信号的脂质体：磷脂酰纤维糖（3）磷酸［PI（3）P］、磷脂酰纤维糖（4，5）焦磷酸［PI（4，5）P2］、磷脂酰纤维糖（4）磷酸［PI（4）P］、磷脂酰纤维糖（3，4）焦磷酸［PI（3，4）P2］和磷脂酰纤维糖［PC］。每一对绿点对应于微阵列上结合于探针的一个蛋白质，每处的点是成双的（结合到蛋白质或脂质体探针）。对应于第二排和第四排阳性点（绿点）的红点被框出。（*Source*：Adapted from Zhu et al.，*Science* 293 (2001) Fig. 2A，p. 2102.）

第二，Tong 及同事用计算机手段在酵母蛋白质组中搜寻同源肽段靶序列。这一步产生了如图 24.31 所示的蛋白质网络关系。之所以是网络关系，是因为很多靶蛋白有它们自己的 SH3 域，反过来又与其他靶蛋白结合。被归于“*k* 核心”（*k*-core）的蛋白质，每一种都会与其他蛋白质产生 *k* 个互作。例如，6-核（6-core）是一组蛋白质，其中每个蛋白质至少能和其他 6 种蛋白质反应。图 24.31a 的黑色区域为 6-核，图 24.31b 是其放大图。

第三步，Tong 及同事用另一种方法——酵母双杂交分析法检测了 SH3 域与靶蛋白的互作。他们比较了两种方法的结果，发现结果是一致的。所有互作中有 59 个能被两种方法同时检出。由于它们独立地由两种方法鉴定出，所以其中绝大多数是可信的。Tong 及同事又做了一个实验，他们选择了一个带有 5 个不同富含脯氨酸域的蛋白质，也证明了其正确性。

每一种检测蛋白质-蛋白质互作的技术都很有用，但是它们都有各自的问题。所有的技术都会产生假阳性（检测到一个明显的相互作用，但是在体

内不会发生）和假阴性（未能发生明显的相互作用）结果。将不同技术相结合可能会产生最好的结果。

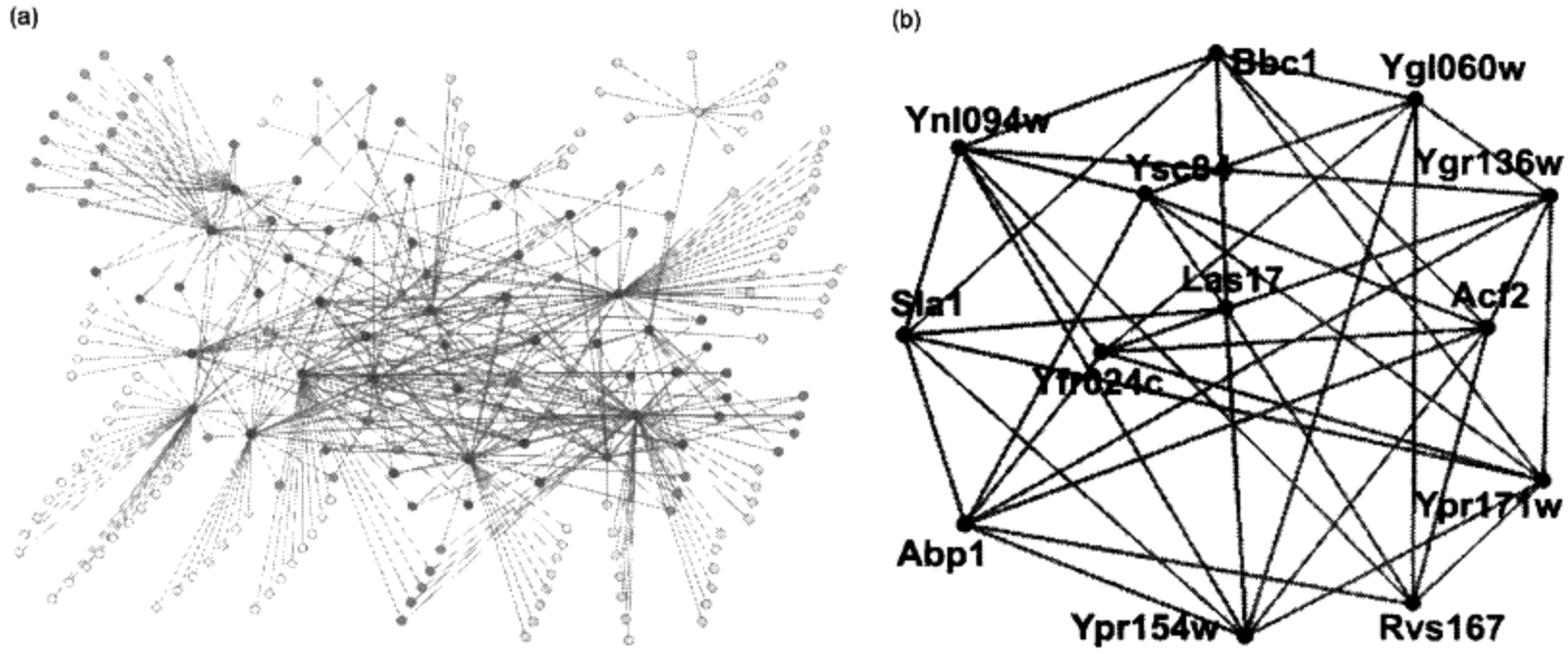


图 24.31　预测的蛋白质-蛋白质互作的网络关系，关于酵母 SH3 结构域及其靶位点。（a）由噬菌体展示技术和酵母蛋白质组搜索技术预测的所有蛋白质和它们之间的互作。蛋白质集中于 *k* 核，其中每个蛋白质都有 *k* 个相互作用。例如，3-核的蛋白质都产生 3 个相互作用。每个蛋白质根据其 *k* 值不同，其颜色如下：6-核，黑色；5-核，蓝绿色；4-核，蓝色；3-核，红色；2-核，绿色；1-核，黄色。6-核蛋白质的互作以红线表示。（b）6-核网络放大图，显示相互作用的具体蛋白质。（*Source*：Adapted from Tong et al., *Science* 295（2002）Fig. 2, p. 322.）

小结　很多蛋白质通过与其他蛋白质互作而表现其功能。检测蛋白质-蛋白质互作的技术有多种。传统的方法是酵母双杂交分析法，目前还有其他几种方法也已应用，包括蛋白质微阵列分析、免疫亲和层析-质谱联用法，还有如噬菌体展示技术与计算机结合的方法。这些分析最有用的成果之一是发现新蛋白质的功能。

24.5　生物信息学

由于数据库塞满了数以十亿计的人和其他生物的基因组碱基序列和不计其数的蛋白质结构及其互作的信息，所以操作和使用这些数据就成了一个关键问题。由此产生一门新的学科——生物信息学。生物信息学的实践者必须兼通生物学和计算机数据处理，以便管理从基因组和蛋白质组研究中得到的数据，然后编写程序供科学家使用。例如，BLAST 是一个用于在数据库中搜寻与你感兴趣的序列相似的 DNA 或蛋白质序列并告知它们的同源性的程序；GRAIL 则是一个从已知数据库中鉴定基因的程序。

人们已建立了两类数据库。第一类是包括所有生物的 DNA 和蛋白质序列的数据库。两个已普及的 DNA 序列数据库是 GenBank（http://www.ncbi.nlm.nih.gov）和 EMBL（http://www.ebi.ac.uk/embl/），Swissprot（http://www.ebi.ac.uk/Swissprot/）是一个普及的蛋白质序列数据库。第二类是为某些特殊生物建立的专门的数据库，例如，FlyBase 是果蝇基因组数据库，你可以访问 http://www.flybase.bio.indiana.ede.82进入该数据库，查询果蝇遗传图谱、基因、DNA 序列和一些其他信息。基因组研究所（The Institute for Genomic，TIGR）有一个网站（http://www.tigr.org）能链接到 TIGR 数据库，包括很多微生物的完整基因组信息。

正如 Willian Gelbart 指出的，问题在于我们对基因组序列的功能还不了解。他打了个比方：我们知道了很多“名词”，即知道了基因组的多肽编码区，但我们还不知道“动词”、“形容词”和“副词”，即不知道基因何时表达及表达量。我们也不知道支配多肽的组装和工作的“语法”，不知道它们是怎样催化生化途径的。生物信息学会提供数据库及全面理解基因组学“语法”所需的注解。

小结　生物信息学的任务是建立和使用生物学数据库，其中的一些包括基因组的 DNA 序列。生物信息学对于从庞大的生物学数据中挖掘有关基因结构和表达的有用知识而言是必需的。

从哺乳动物基因组中发现调控基序

以下是科学家以纯计算机工具在哺乳动物基因组中发现调控基序的例子。在研究的发现阶段，这

些科学家既没有用试管［体外（*in vitro*）研究］也没有用细胞或动物［体内（*in vivo*）研究］。因此，他们的工作可以描述为**计算机模拟**（in silico），与他们计算机中的以硅为基础的芯片有关。

在本章前面，我们曾看到一个通过实验鉴定已知转录因子靶位点的例子，但是如何处理尚未鉴定出作用分子的调控位点呢？在 2005 年，Eric Lander、Manolis Kellis 及同事报道了他们用生物信息学解决这一问题的结果，推测调控基序（6～10bp 长）最可能在基因的上游调控区（转录因子可能结合的位点）和 3′-UTR（miRNA 和其他调控分子结合的位点，调控 mRNA 的稳定性和可译性）中发现。他们进一步推测有亲缘关系生物体的调控基序可能是保守的，因此比较了人类、小鼠、大鼠和狗的基因组以发现在 5′侧翼区和 3′-UTR 区中的保守序列。

他们将研究集中于在这 4 个物种中均被详细注释的 17 000 个基因，将每个基因的启动子区限定在以转录起始位点为中心的 4kb 区域内，并将 3′-UTR 限定在每个 mRNA 所注释的翻译终止密码子和 poly（A）信号之间的区域。作为对照，他们参照许多基因最后两个内含子的约 123Mb 的序列，通常认为末端的内含子缺乏调控基序，因此可作为很好的阴性对照。

他们将“保守”定义为：在所有 4 个物种中绝对保守的基序。“保守率”（conserved rate）是一个基序保守次数与其在研究涉及人类基因组区域（如启动子区）内出现次数的比率。最后，“基序保守值”（motif conserved score，MCS）是一个基序的保守率超过同样大小的随机基序的保守率而获得的标准差值。

为阐明保守情况，选择了 8mer（TGACCTTG）的基序（为转录因子 Err-α 的结合位点）。该基序在人类基因组中出现 434 次，其中 162 处为保守，因此，保守率是 162/434（37%），而启动子区随机的 8mer 序列保守率仅为 6.8%。而且，这个 8mer 基序的保守是启动子区特异性的，在内含子区，其保守率仅为 6.2%。根据以上和其他的一些数据的统计分析计算保守值。Err-α 基序的保守值是 25.2 标准差，反映了以 6.8%比率为背景发现 37%保守率的可能性。

为了获得有关调控基序保守性的一般情况，他们计算了 TRENSFAC 数据库中已知转录因子结合位点的 MCS。发现其中 63%基序的 MCS＞3，且 50%基序的 MCS＜6。因此，限定“高度保守基序”的 MCS＞6。作者给出 3 个原因解释为什么许多已知调控基序 MCS 还不到 3，即这些基序鉴定有误，基序在所研究的 4 个物种中不保守，这些基序还不够普遍。

他们在启动子区鉴定了 174 个高保守的基序，其中 59 个与 TRENSFAC 中已鉴定的调控基序严格匹配，10 个弱匹配，另外 105 个基序可能代表了新的调控基序。如果这些新的基序真是调控元件，它们相关的基因很可能在某种程度上呈现组织特异性表达，这是因为同一转录因子控制的基因在同一组织中呈激活状态。作者参考数据库列出了 75 种组织中表达基因的清单，发现已知基序的 86%和新基序的 50%与在一种或多种组织中活性明显增强的基因相关。

检查其真实性的另一个方法是观察元件的位置是否偏向转录起始点。实际上，高保守的元件明显趋向于集中在起始点 100bp 范围内，而随意元件随机分布于分析所涉及的 4kb 的整个区域。总之，这些数据说明鉴定的多数基序很可能属于真正的调控元件。

在基因的 3′-UTR 发现了 106 个高保守的基序，由于没有类似于 TRENSFAC 的 3′-UTR 元件数据库，所以必须用其他的方法检测其真实性。幸运的是，有两个很突出的特征可以被我们利用。首先不像启动子区的元件，3′-UTR 内的基序呈现明显的方向偏向，即倾向存在于一条链上而不在另一条链。这与 3′-UTR 基序的活性存在于 mRNA 上的假说是一致的，这些基序在此处与 miRNA 或其他分子结合并调控 mRNA 的稳定性和可译性。这是因为这些基序必须在被转录为 mRNA 的那条链上。其次启动子区基序的活性体现在 DNA 水平，它们典型地结合激活因子，并在两个方向均呈现活性，因此不存在链的偏向性。

基因 3′-UTR 基序的第二个特征是明显偏向于长度为 8 个碱基，并且 A 在最后一位。启动子区的基序不存在这一偏好。这一特征与以下假说一致，即此 8mer 序列是 miRNA 杂交位点的长度，miRNA 倾向于以 T 起始，并跟随与受调控 mRNA 相应位点互补的 7 个碱基。

高保守的 8mer 和 miRNA 之间的关系令人感兴趣，因此搜索了 miRNA 的注册库（包含 207 个不同的人类 miRNA）并将其与这些 8mer 序列进行比对，结果发现已知人类 miRNA 的 43%与这些 8mer 序列中的一个完美匹配，而只有 2%与对照组的 8mer 序列相匹配。不与已知 miRNA 相匹配的 8mer 序列比相匹配的进化的更快。说明匹配的 8mer 序列没有破坏与 miRNA 的杂交，故不能改变它们的序列，这对基因的调控很重要。

最后，可用保守的 8mer 基序去发现新的基因。从 4 个基因组中寻找与高保守 8mer 基序互补的保守

序列。继而检测保守序列周围的序列形成稳定茎环结构（miRNA 的特征）的能力，发现 242 个这样的稳定茎环结构，可能编码 miRNA。其中 113 个编码了已知的 miRNA，剩余 129 个的编码产物预计也是 miRNA。随机选择了其中的 12 个，并检测它们在成人组织库中的表达（此项工作中唯一的体外试验）。发现有 6 个基因表达，因此，129 个预测的 miRNA 基因中有许多的确编码了 miRNA。这意味着尚有许多 miRNA 基因有待发现，且 miRNA 对基因表达的调控可能比我们预期的更为广泛。

小结 利用计算机生物技术，Lander 和 Kellis 在包括人类在内的 4 种哺乳动物的启动子区和 3′-UTR 中发现了高保守基序。在启动子区的基序可能代表转录因子的结合位点。许多 3′-UTR 内的基序可能代表 miRNA 结合位点。

如何使用数据库

一些非常有用的数据库存储在国家生物信息学中心（National Center for Biological Information，NCBI）。在这一节我们将举一些如何获得并利用这些数据的例子。为了搞清楚怎样进行搜索，假设你是一个内科医生，正在给一个长疣的患者看病。你怀疑某种病毒是病因，甚至你的大脑中已经想到了一个待检病毒（乳头状瘤病毒），你会将疣切下放在含有去污剂的缓冲液中均质化，使之释放病毒颗粒，纯化 DNA，再用设计的待检病毒特异性引物进行 PCR。如果得到一条特异性条带，说明待检的病毒 DNA 在疣中存在。为了进一步研究这个病毒确切的株系，可以对扩增产物进行测序，得到如图 5.19 所示的测序胶。

如果你需要练习解读测序胶，那就不要看下面所示的序列，直接写出前 21 个碱基，以胶最上面的 C 打头。如果你想跳过这一步，请看下面，这就是它的序列：caaaaaacggaccgggtgtac。小写的目的是防止混淆，小写 c 和 g 比大写 C 和 G 容易分辨。

进入 NCBI 主页点击 BLAST 开始搜索，或者直接从 NCBI BLAST 的主页（http://www.ncbi.nlm.nib.gov/BLAST/）开始。如果你要在核酸数据库中搜索，向下寻找标题“Nucleotide”并点击“Nucleotide-nucleotide BLAST”（BLAST 发音为 blastn）。标有“Search”的框将带你寻找你想查的序列（或想在数据库中比对的序列）。你可以用键盘输入序列，但是直接从文档中拷贝并粘贴会更容易些。

当已输入了序列，用“Set subsequence”框可以选择查找序列的一部分。例如，如果想搜索序列的第 10 到第 21 个残基，请在“From”窗口中输入 10，在“To”窗口中输入 21。也可用“Choose database”选择数据库。默认的数据库是 nr，代表“nonredundant”。这个数据库包括几个数据库的所有核酸序列，是最全面的。开始搜索请点击序列框下的“BLAST!”，之后会出现一个对话框，告诉你序列输入成功并预测完成搜索的大概时间。这个过程有时仅需几秒钟，但是通常需几分钟甚至几小时。请按指令点击“Format!”即可获取结果，同时你会得到一个请求 ID（request ID，RID）号。

如果立刻得到了结果，你就可以进行分析了。但也许得等待。在这种情况下，记住你的 RID 号，你可以再过一段时间登录 NCBI 获取这个 RID 号的搜索结果。怎么做呢？请看 BLAST 结果页上端的附近，在标识语“NCBI nucleotide-nucleotide BLAST”下面，点击“Retrieve result of an RID”。计算机会提示你输入 RID 号。输入之后，如果有结果，就会立刻显示出来。

所得的结果会有几种形式。第一，你会看到一张彩色的条形图，这张图说明了查询序列和数据库中不同序列的粗略匹配程度。红色代表最好的匹配。可以将鼠标指在每个条上读出两者的一致性（identity）。在条形图的下面是与查询序列在一定范围匹配的序列，这个问题后面还会讨论。对每一匹配序列已进行了一致性分析，匹配程度最高的放在第一，接着按匹配的紧密程度按递减顺序排下一个。

NCBI 会给每一对序列打两个分数。第一个是**“比特值”**（bit score，*S*），它与查询序列和数据库序列间匹配数有关。分值越大，匹配越好。本例中，最好匹配的 *S* 值是 42.1 分。第二个是**“期望值”**（expect value，***E* value**），表示预期随机看到产生相应比特值的匹配数。所以，*E* 值越小，匹配越好。匹配很好时期望值远小于 1.0。在这个实例中，最佳匹配的期望值是 0.011。

那么，数据库中与查询序列之间最佳匹配序列的身份（identity）是什么？将鼠标置于顶部的条形（蓝色）上，显示它属于人乳头状瘤病毒（HPV）31 型，表明患者的疣由这种病毒引起。你可以从所有的条形下面罗列的配对序列得到相同的信息。两个黑条对应的是小家鼠（*Mus musculus*）的基因，匹配得也很好。

随着从上往下查看搜索结果，可以看到查询序列和数据库序列比对的结果。你会看到查询序列与 HPV31 匹配得最完美，而和小鼠的 21 个位点仅有 19 处匹配。虽然表面上看起来差异很小，但是 *E* 值差异却很大。比对中小鼠基因的 *E* 值为 0.17，比 0.011 大得多。

输入的查询序列仅有 21 个碱基，非常短。增加序列的长度有什么影响呢？要么再读一段测序胶

(至 42 个碱基)，或者输入以下序列重新查询：

caaaaaacggaccgggtgtacaacttttactatggcgtgagc

这次的 E 值是 e^{-14}，非常小的一个数，说明 42 个位点的完美匹配有更高的生物学意义。

如果你没有自己的 DNA 序列，但又想利用 NCBI 数据库获得一般性信息，该怎么办呢？例如，你可能有兴趣找一些与人类疾病相关的基因，如结肠癌。要进行这种搜索，可以访问 NCBI 点击“Genomic Resources”，然后点击“Human Genome Resources”，再点击“Maps and Markers”下的“Map Viewer”，大约在页面的中下部将显示一张人类染色体组型示意图，展示了人的所有染色体，包括线粒体染色体（mitochondrial chromosome，MT）。如果你想了解结肠癌基因，但不知道查找哪条染色体，只需在“Search for”下键入“colon cancer”（结肠癌），另外两格空着不填。然后点击右边的“Find”键。系统会突出强调除 4 条染色体之外其他染色体上的许多基因（至少 2006 年 11 月是这样，随着人们对结肠癌基因的进一步了解，这个数还会变化）。点击 3 号染色体下的蓝色数字“3”，会看到一个标有 *MLH1* 的基因，边上还带了一个红色标记。

点击 *MLH1* 基因链接到这个基因的数据。概要会告诉你 *MLH1* 基因是 *E. coli mutL* 基因在人类的同源基因，编码一个参与错配修复的蛋白质。人的 *MLH1* 基因也参与错配修复。它的突变会造成错配和突变积累，这可能是人患癌症的前奏，结肠癌尤其如此，因为在正常情况下阻止癌症发生的基因［**抑癌基因**（tumor suppressor gene)］会由于突变而失活，使细胞生长抑制基因［**原癌基因**（oncogene)］会由于突变而被激活。

也可以用 NCBI 了解到蛋白质结构的信息。例如，想了解 p53 蛋白（一种抑癌基因产物），其失活是很多癌症的特征。可访问 NCBI 并点击“Structure”。下一页会显示“Search Entrez Structure for…”，在空格处键入“p53”并点击“Go”，会得到一个条目表。到条目的第 5 页点击链接名为“1TUP”的 90 号条目（至 2006 年 11 月）名为“与 DNA 复合的肿瘤抑制物 p53”。点击“View 3D Structure”，服务器会提示你下载 CN3D 浏览器。如果没有，点击“GetCn3D4.1”链接，根据提示可以免费下载并安装这种软件。

一旦安装了 CN3D 浏览器，再回到上一步点击“View 3D Structure”，会看到 p53 与 DNA 形成的复合物的 X 射线晶体衍射结构图。有了 CN3D 软件，还可以用鼠标根据你的意愿旋转这个结构图。用鼠标按住结构图的左边，再将鼠标拖至右边，结构图就会从左至右旋转。也可以从上向下旋转，在水平和垂直方向之间以任意角度旋转。尽可旋转，直到看清 DNA 大沟和锌指结构的相互作用为止。

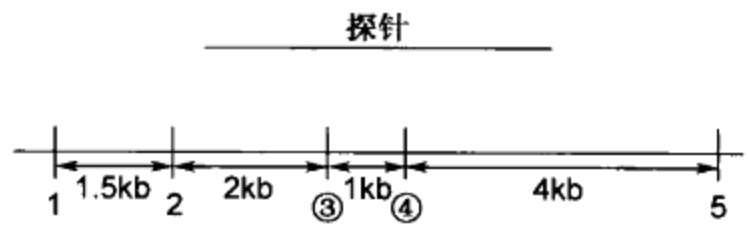
可以查阅以前章节的部分蛋白质的 3D 结构。例如，查阅 GAL4 的结构和糖皮质激素受体。这两种结构图都可以旋转，比本书中清楚得多。

小结 NCBI 包含了一个巨大的生物信息库，包括基因组和蛋白质组数据。以某段序列为出发点，可以发现该序列所属的基因，并将其与相似的基因比较；也可以以某个想研究的话题为出发点搜索数据库得到关于此话题的信息；还可以查阅蛋白质，通过在屏幕上的旋转观察其三维结构。

问题解答

问题 1

下图是用 RFLP 标记的一段 DNA 物理图谱。



标有号码的垂线代表 *Pst* I 的限制酶位点。位点 3 和 4 具有多态性，其他没有。*Pst* I 限制酶位点的间距已给出。检测 RFLP 的探针范围用图上的水平线表示。将不同个体的 DNA 用 *Pst* I 酶切，片段进行电泳、转膜及与带标记的探针杂交。

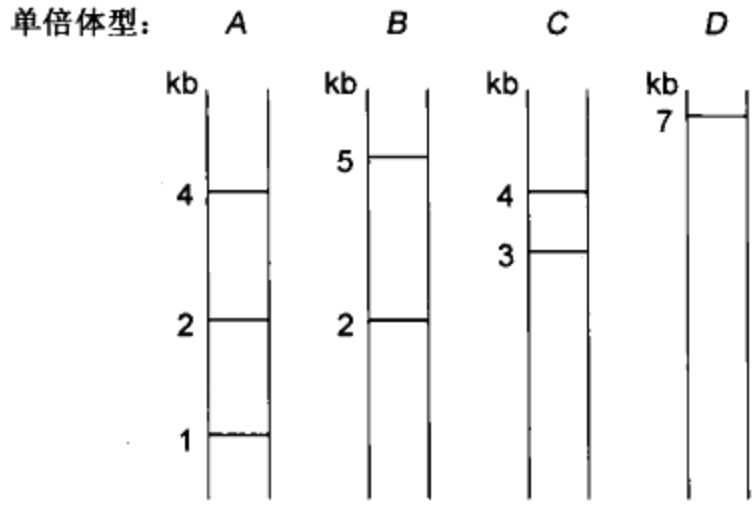
a. 如果 DNA 来自以下单倍体型所形成的纯合子个体，请绘出将会看到的印迹图。

单倍体型	位点 3	位点 4
A	存在	存在
B	存在	不存在
C	不存在	存在
D	不存在	不存在

b. 位点 1 存在与否对结果又何影响？

解答

a. 样本 A：如果位点 3 和 4 都存在，则 4 个片段都产生。从左至右分别为：1.5kb、2kb、3kb、4kb。当然，2kb、3kb、4kb 的片段会与探针杂交显出条带。样本 B：位点 3 存在，所以将产生 2kb 的片段；位点 4 不存在，所以 1kb 的片段会与 4kb 的片段融合产生一个 5kb 的片段。样本 C：位点 3 不存在，2kb 的片段与 1kb 的片段融合产生一个 3kb 的片段；位点 4 存在，产生 4kb 的片段。样本 D：位点 3 和 4 都不存在，所以 2kb、1kb、4kb 的片段会融合在一起，产生一个 7kb 的片段。下图显示你得到的结果。



b. 位点 1 的存在与否对结果没有影响，因为探针的范围没有达到位点 2 的左边。所以，即便位点 1 消失，产生一个位于 2 左边的较长片段，用该探针也无法检测到这个片段。

问题 2 假如你对一个导致人类遗传病的基因作图。用 B-101 探针检测，发现这个基因与一个 RFLP 连锁。将带标记的 B-101 与鼠人杂种细胞的 DNA 杂交。下表显示存在于每个杂种细胞系的人类染色体组以及探针与其杂交的情况。

细胞系	包含的人类染色体	探针杂交
A	1，2，17	−
B	3，5，9，12	+
C	5，18，22	−
D	1，3，9	+
E	2，9，18	−
F	2，3，7	+

请问人的哪个染色体携带致病基因？

解答

请看每个阳性细胞系（B、D 和 F）包含的人类染色体组。它们是：

B：3、5、9、12；

D：1、3、9；

F：2、3、7。

现在发现有一个染色体在三者中均存在，它就是 3 号染色体。所以，RFLP 定位于 3 号染色体，致病基因与 RFLP 连锁。

总结

有几种方法可以从尚未测序的大段 DNA 区域中鉴定基因。其中之一是外显子捕获法，采用一种特殊的载体将外显子单独克隆出来。另一种是用甲基化敏感的限制性内切核酸酶寻找 CpG 岛——DNA 中包含非甲基化 CpG 序列的区域。在基因组时代之前，遗传学家用图位法将亨廷顿症基因（*HD*）定位在 4 号染色体末端附近的区域。然后他们又用外显子捕获法鉴定了该基因。

很快，DNA 自动测序法使生物学家能够得到病毒和生物体的碱基序列，包括从简单的噬菌体到细菌、真菌、简单动物、植物、小鼠和人类。人类基因组计划的很多绘图工作是靠酵母人工染色体（YAC）完成的，YAC 是一种包括酵母复制区的染色体，包括一个着丝粒、两个端粒。长达 1Mbp 的外源片段能够插入到着丝粒和一个端粒之间。另外，由于细菌人工染色体（BAC）极高的稳定性和易操作性，绝大多数的测序工作靠它来完成。BAC 是一种基于大肠杆菌 F 质粒的载体，最大可插入 300kb 左右的片段，但平均为 150kb。

对人类基因组或任何大的基因组作图，需要一套路标（标记）以确定基因的位置与标记间的关系。基因本身可以作为标记，但是常用的标记是一些未知的非基因 DNA 片段，如 RFLP、VNTR、STS（包括 ETS）和微卫星。RFLP（限制酶切片段长度多态性）指的是用一种限制性内切核酸酶对两个或多个个体的 DNA 进行完全酶切后产生的限制性片段长度的不同。RFLP 由限制性酶切位点在某处的有无，或两酶切位点之间碱基的插入或缺失决定。RFLP 也可由两酶切位点之间的串联重复序列的重复次数不同（VNTR）决定。STS（序列标签位点）是指通过一对引物进行 PCR 扩增获得预计大小 DNA 片段来鉴定的 DNA 区域。EST（表达序列标签）是 STS 的一种，由 cDNA 产生，因此它们代表表达的基因。微卫星是由 PCR 产生的一种 STS，PCR 引物对位于 2～4nt 串联重复序列的两侧。

当 STS 和其他标记相隔甚远以至于不在同一个 BAC 上时可用辐射杂种作图。在辐射杂种作图法中，人细胞受辐射而致染色体断裂，这些逐渐死亡的人类细胞与仓鼠细胞融合。每个融合细胞含有一组不同的人类染色体片段。两个标记挨得越近，它们越容易在同一杂种细胞中被发现。

大规模的测序工程可以有两种形式：一种是先作图再测序（逐步克隆测序）；另一种是鸟枪法。实际上，这些方法的结合已经用于人类基因组测序。逐步克隆测序要求先得到包括 STS 的物理图谱，然后对重叠克隆群（多个 BAC）测序，通过这些重叠序列可以将所有序列拼凑起来。鸟枪法要求构建一个含有不同大小插入片段的克隆文库，然后随机测定这些插入片段的序列，这种方法必须依赖于计算机程序去发现序列之间的重叠区，然后把它们拼接在一起。

对人类 22 号染色体长臂的序列测定揭示出下列内容。

1. 已有的方法尚不能填补空缺。

2.679 个基因被注释。

3. 大部分（约 97%）的染色体由非编码 DNA 组成。

4.40%的染色体是零散分布的重复序列，如 Alu 序列和 LINES。

5. 染色体上不同区域重组率变化很大，低重组率的长区域被高重组率的短区域所间断。

6. 包含一些局部和长距离间隔重复序列。

7. 较大区域中基因间连锁的区段在小鼠的 7 个不同染色体中都能找到相应的同源区。

人类染色体 21q 和 21p 的一小部分已经被测序，留下了相对较少且较短的间隙。这些序列揭示该染色体基因较少，仅约 225 个，另有 59 个假基因。小鼠 10 号染色体和人 21 号染色体共有的 24 已知基因以相同的顺序出现在两个染色体中。

人类 X 染色体 151Mb 序列（常染色质的 99.3%）揭示了 1098 个蛋白质编码基因（包括 168 个与性连锁表型相关的基因）和 173 个非编码 RNA 的基因。X 染色体富含 LINE1 元件，其在女性 X 染色体失活机制中可能扮演了驿站的角色。人类 X 染色体序列和鸡基因组序列之间的比较结果证实了 X 染色体（以及 Y 染色体）源于一对常染色体。3 种哺乳动物 X 染色体序列的比对证明这些染色体保持着高度保守的同线性（synteny），可能反映了进化压力稳定地保持了这些基因在 X 染色体上的顺序。

人类基因组草图已由两个独立的研究组报道，结果显示人类基因组可能比预计含有更少的基因。基因组的一半来自于转座子的作用，转座子自身为基因组贡献了许多基因。除此之外，细菌似乎也贡献了不少基因。已完成的人类基因组草图比工作草图更加完善和准确，但它仍然存在一些缺口。基于最终的草图，遗传学家估计基因组包含 20 000～25 000个基因。最终的草图也在人类进化过程中基因的生死方面提供了有价值的信息。

比较人类和其他脊椎动物的基因组，发现了很多相似和不同之处。这些比较还有助于鉴定一些人类基因，而且也将有助于找出人类遗传病的致病基因。我们还可以利用亲缘关系相近的物种，如通过研究小鼠基因的表达规律，来预测人类相应基因的表达规律。小鼠和人类染色体的详细比较表明，这两个物种之间有很高的相似性。

通过每次突变一个基因来研究哪些基因是生命必需的，并由此定义简单生物体的必需基因组是可能的。原则上讲，此方法定义最小基因组——生命所必需的最小组基因也是可能的。有可能最小基因组大于必需基因组。理论上讲，将最小基因组转到缺失自身基因的细胞中并创造新的生命形式也是可能的，这种新的生命形式可以在实验条件下存活和繁殖。添加选择性基因后，这种被改造后的生命形式可以行使多种有用的功能。

通过建立条形码来识别地球上所有生物体的行动已经开始，第一个“生命条形码”将由每个生物体都存在的线粒体 COI 基因中一段 648bp 的序列组成，此序列足以特异性地鉴定出几乎所有的动物。另一个序列，或称条形码，是为植物而设计，也许最终可以延伸到微生物。

功能基因组学是对大批量基因表达的研究。该领域的一个分支是转录组学，即转录组的研究，转录组是一个生物体在任一给定时间产生的所有转录物。转录组学研究的方法之一是创建 DNA 微阵列或微芯片，其上分别载有成千上万的 cDNA 或寡聚核苷酸，然后与标记的 RNA 或相应的 cDNA 杂交。每一个点杂交后的荧光信号强度揭示出对应基因的表达水平。这种分析可以一次性分析许多基因的时空表达。

SAGE（基因表达的系列分析）能够测定组织中哪些基因得以表达以及这些基因的表达程度。特定基因的短标签产生于 cDNA 并通过接头连接在一起。对连接的标签进行测序即可确定哪些基因得以表达及其表达量。

高密度全染色体转录制图研究证明，细胞质 poly（A）$^+$ RNA 的大部分序列源于 10 条人类染色体非外显子区。而且这 10 条染色体的近一半转录产物是 poly（A）$^-$ 转录物。总的来说，这些结果表明 10 条染色体在细胞质和细胞核内的大部分稳定的转录物源于外显子以外的区域。这可能有助于解释物种间的巨大差异，如人和黑猩猩的外显子几乎是相同的。

通过创建突变体可进行基因组功能分析，即利用抗生素基因一次替换一个基因，并在一侧连接一段寡聚核苷酸作为分子条形码以识别突变体。然后将这些突变体混合培养在各种不同的条件下以观察各种突变体消失的速度。利用 RNAi 失活基因也可进行基因组功能分析。

通过检测被外源性 miRNA 消减的 mRNA 谱并将其与各种组织中在 mRNA 水平上的基因表达谱比较，可以开展组织特异性表达分析的研究。假如一 miRNA 引起该 miRNA 表达细胞中天然表达较低的 mRNA 的表达降低，则提示这些 miRNA 至少部分地导致这些 mRNA 的天然低表达，这类分析暗示 miR-124 破坏了脑组织中一些 mRNA 的稳定性，而 miR-1 破坏了肌肉组织中一些 mRNA 的稳定性。

染色质免疫沉淀后的 DNA 微阵列分析能够鉴定激活因子和其他蛋白质的 DNA 结合位点。对基

因组较小的生物体（如酵母）而言，微阵列分析包括了所有基因间的区域。但对大基因组而言（如人类基因组）这无法实施。CpG 岛可用于缩小微阵列分析的范围，因为它与基因的调控区域相关。如果一个激活因子的活性作用时间和条件已经知道，则在这些时间或条件下处于激活状态的基因的调控区也可用于此目的。

小鼠可以作为人类大规模基因表达研究的替代品，因为就伦理而言，不可能用人做这方面的研究。例如，科学家已经研究了小鼠几乎所有的和人 21 号染色体上同源的基因的表达，并且追踪了这些基因在胚胎发育不同阶段的表达，并且已经列出了表达这些基因的胚胎组织名录。

单核苷酸多态性能够引起许多由单基因甚至多基因控制的遗传疾病。它们也能用来预测人对药物的反应。含一百多万个 SNP 的单倍体型图谱使我们很容易从没有影响的 SNP 中挑选出重要的 SNP。结构变异（缺失、插入、倒位和完整的大段 DNA 的重排）也是人类基因组变异的很重要的来源。理论上一些结构变异可能会使某些人更易患病，有些规则是良性的，而有一些却是有益的。

一个生物的所有蛋白质产物的总和称为蛋白质组，研究这些蛋白质或其中很小的一部分即为蛋白质组学。当前的蛋白质组学研究首先要求分离蛋白质，有时是大规模的分离。对大量蛋白质进行同时分离的最好工具是双向电泳（2-D）。蛋白质分离之后必须鉴定，最好的方法是用蛋白酶切割成肽段，再用质谱分析鉴定这些肽段。有朝一日，带有抗体的蛋白质微芯片可能会取代分离法而分析复杂的蛋白质混合物。

很多蛋白质通过与其他蛋白质互作而表现其功能。检测蛋白质-蛋白质互作的技术有多种。传统的方法是酵母双杂交分析法，目前还有其他几种方法也已应用。包括蛋白质微阵列分析、免疫亲和层析-质谱法，还有如噬菌体展示技术与计算机结合的方法。这些分析最有用的成果之一是发现新蛋白质的功能。

生物信息学的任务是建立和使用生物学数据库，其中最复杂的包括基因组的 DNA 序列。生物信息学对于从庞大的生物学数据中挖掘有关基因结构和表达的有用知识而言是必需的。

利用计算机生物技术，Lander 和 Kellis 在包括人类在内的 4 种哺乳动物的启动子区和 3′-UTR 中发现了高度保守基序。在启动子区的基序可能代表转录因子的结合位点。许多 3′-UTR 内的基序可能代表 miRNA 结合位点。

NCBI 包含了一个巨大的生物信息库，包括基因组和蛋白质组数据。可以以某段序列为出发点，发现该序列所属的基因，并将其与相似的基因比较；也可以以某个想研究的话题为出发点，搜索数据库得到相关的信息；还可以查阅一个蛋白质，通过在屏幕上的旋转观察其三维结构。

复习题

1. 何谓 CpG 岛？为什么人类基因组的 CpG 序列趋于消失？

2. 哪类基因突变会引起亨廷顿症（HD）？能证明 *HD* 基因确实为亨廷顿症的证据是什么？

3. 请用一个假说来解释亨廷顿症中一段延伸的多聚谷氨酰胺（extended polyglutamine tract）的神经毒效应。给出证据支持你的假说。

4. 何谓可读框（ORF）？写出一段含有短的 ORF 的 DNA 序列。

5. YAC 载体中的必要元件有哪些？

6. BAC 载体是以哪种质粒为基础的？它们的必要元件有哪些？

7. 描述在基因组中寻找 STS 的步骤。

8. 阐述微卫星和小卫星。在绘制连锁图中，为何微卫星法优于小卫星法？

9. 在一组 BAC 克隆中怎样利用 STS 构建一个重叠群（contig）？绘图说明（与书中例题不同）。

10. 描述辐射杂种法在 STS 定位中的应用。

11. 表达序列标签（EST）与普通的 STS 有何不同？

12. 比较和对照逐步克隆法与鸟枪法对较大基因组进行的测序。

13. 从人类 22 号染色体的序列中我们可以得出哪些主要结论？

14. 从人类 21 号染色体的序列中我们可以得出哪些主要结论？

15. 东方红鳍鲀的基因组是人类基因组大小的 1/9，但包括的基因数量却差不多，怎么解释这种现象？

16. 在人类基因组和小鼠基因组中的“同线性区域”（syntenic region）指什么？

17. 生物的“必需基因簇”与“最小基因组”有何不同？

18. 请比较“结构基因组学”和“功能基因组学”。

19. 请阐述制备 DNA 芯片（寡聚核苷酸阵列）的过程。

20. 设计一个实验，用 DNA 微阵列检测病毒感染细胞后两个不同时期的基因转录物。给出实验结果。

21. 描述如何用 SAGE 实验检测某种癌细胞的

转录情况。解释双标签如何形成，并以你获得的实际序列来说明。

22. 何谓 SNP？为何大部分 SNP 不重要？某些 SNP 怎样才是有用的？它们是怎样被滥用的？

23. 请对比转录基因组学和蛋白质组学在所使用技术和获得信息方面的异同。

24. 描述利用生物信息学方法鉴定人类调控基序。

分析题

1. 以下所列 DNA 片段能否用外显子捕获法检测？请给出原因。

a. 内含子；

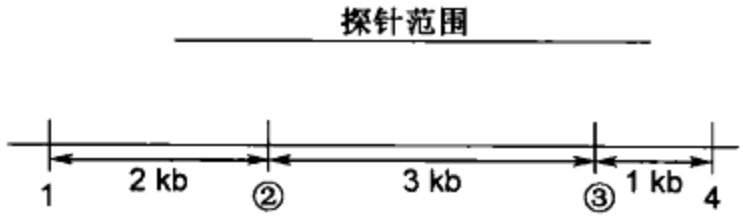
b. 外显子的一部分；

c. 两端都有一部分内含子的一个外显子；

d. 仅在一端有一部分内含子的一个外显子。

2. 请按顺序列出制备含有二核苷酸 AT 和 AC 的寡聚核苷酸阵列的步骤。可忽略除这两个点之外的所有点。

3. 下图是一张由 RFLP 分析而绘出的一段序列范围的物理图谱。



下面标有号码的垂线代表 *Sma* I 限制酶位点。带圈的位点 2 和 3 具有多态性，其他位点不具多态性。酶切位点的间距已给出。用 *Sma* I 对 DNA 片段进行酶切、电泳、转膜再与标记的探针（图的顶部示意了探针的范围）杂交。如果以下列单倍体型所形成的纯合子为例，请给出针对位点 2 和 3 所能检测到的条带大小。

单倍体型	位点 2	位点 3
A	存在	存在
B	存在	不存在
C	不存在	存在
D	不存在	不存在

4. 你正在对一个导致人类遗传病的基因作图。发现这个基因与一个用 X-21 探针检测到的 RFLP 连锁。将标记的 X-21 序列与一系列鼠-人杂交细胞的 DNA 杂交。下表所示的是存在于每个杂交细胞系的人类染色体组，以及探针与其杂交情况。请问人的哪个染色体携带致病基因？

细胞系	包含的人类染色体	探针杂交
A	1，5，21	−
B	6，7	+
C	1，22，Y	−
D	4，5，18，21	+
E	8，21，Y	−
F	2，5，6	+

5. 请用 BLAST 搜索以下序列（此序列以 10 个核苷酸为一组分开）的前 20 个核苷酸：ttaagtgaaa taaagaagtga atgaaaaaat aatatcctta。你鉴定出是什么基因？最优 *E* 值是多少？然后再用所有 40 个核苷酸试试。这次你仍然得到同样的基因吗？这次最优 *E* 值是多少？为何这次 *E* 值会不同？这个基因定位在哪条染色体？此基因与男性前列腺癌有没有关系？如果有，是什么关系？

翻译　马正海　校译　马　纪

分子生物学专业词汇表

A

AAUAAA：动物多聚腺苷酸化信号的重要部分，指示在其下游约 20nt 处发生剪切和多聚腺苷酸化。

A site（ribosomal）：A 位点（核糖体上）。核糖体上新的氨酰-tRNA 的结合位点（第一个氨酰-tRNA 除外）。

A site（RNA polymerase）：A 位点（RNA 聚合酶上）。在磷酸二酯键形成过程中引入核苷酸的占据位点。

abortive transcript：中的转录物。在启动子清除之前，原核生物启动子中合成的很短（约 6 个核苷酸长）的转录产物。

Ac：一种玉米转座子（“激活因子”），它提供必需的转座酶激活非自主性转座子 *Ds* 发生转座。

acceptor stem：接受茎。tRNA 分子的 5′端与 3′端碱基配对形成二级结构的一部分。由 3′端通过负载“接受”一个氨基酸而得名。

acidic domain：酸性结构域。富含酸性氨基酸的转录激活结构域。

aconitase：顺乌头酸酶。此酶的脱辅基蛋白（铁离子缺乏时）与 mRNA 上已结合的铁离子响应元件发生互作，调控其翻译或降解。

activation region I（ARI）：激活区 I（简称 ARI）。CAP 的一个区域，被认为与 *E. coli* RNA 聚合酶的 α 亚基羧基末端域（CTD）相结合。

activation region II（AR II）：激活区 II（简称 ARII）。CAP 的一个区域，被认为与 *E. coli* RNA 聚合酶的 α 亚基氨基末端域（NTD）相结合。

activation site：活化位点。氨酰-tRNA 合成酶上的一个位点，使氨基酸活化形成氨酰腺苷酸。

activator：激活因子。与增强子（或激活因子结合区）结合的蛋白质并激活附近启动子的转录。在真核生物中，它激发转录前起始复合体的形成。

activator interference：激活因子干扰。见 squelching。

activator-binding site：激活因子结合位点。原核生物转录激活因子结合的 DNA 位点（如在一种可被代谢产物抑制的操纵子中的 CAP-cAMP 结合位点）。

ADAR：见 adenosine deaminase acting on RNA。

adenine DNA glycosylase：腺嘌呤 DNA 糖基化酶。细菌中称为 MutY，人类中为 hMYH。它能移除与 oxoG 错配的 A。

adenosine deaminase acting on RNA（ADAR）：作用于 RNA 的腺苷脱氨酶。一种 RNA 编辑酶，它使 RNA 中的某些腺苷脱氨基后转变成次黄嘌呤核苷。

A-DNA：在相对湿度较低时发现的一种 DNA 形式，每圈螺旋含有 11 个碱基对。RNA-DNA 的杂合链在溶液中呈现此种形式。

affinity chromatography：亲和层析。基于待纯化分子与固定在树脂上的诱饵分子之间的亲合性来纯化分子的层析方法。诱饵分子可以是抗体、酶底物，或其他任何与待纯化分子有已知的特定的亲和性。

affinity labeling：亲和标记。通过共价结合方式，利用与某种物质具有特异亲合性的反应活性物来标记这种物质（如通过连接一个与某种酶共价结合的底物类似物来标记这种酶的活性位点）。

A1492 and A1493：A1492 和 A1493。16S rRNA 中的两个高度保守的碱基。通过插入到密码子与反密码子配对后形成的小沟里，使两者间的作用趋于稳定，因而在密码子-反密码子的相互识别中起关键作用。

Ago1：见 Argonaute1。

Ago2：见 Argonaute2。

alarmone：预警子。有机体在逆境响应过程中产生的物质；启动逆境应答反应。

alkylation：烷基化。将含碳基团添加到其他分子上的过程。碱基烷基化会产生 DNA 损伤，进而可能导致突变。

allele-specific RNAi：等位基因特异的 RNA 干扰。导入一个经过修改的基因拷贝，其 mRNA 不会被针对该基因内源拷贝的 RNAi 所破坏，从而可以对该修改的基因进行遗传操作，检测在内源基因不表达的背景下的遗传效应。

allolactose：异乳糖。一种重排的以 β-1，6-糖苷键连接的乳糖形式；乳糖操纵子的诱导物。

allosteric protein：变构蛋白。内部一个位点上结合一个分子后改变与之相距较远的位点的构象，从而改变了该位点与第二个分子间互作的一类蛋白质。

alternative splicing：选择性剪接。以两种或更多的

方式剪接相同的前体 RNA，产生两种或多种不同的 mRNA，进而生成两种或更多种不同的蛋白质产物。

Alu element：Alu 元件。人类的一种非自主性反转录转座子，其中含能够被限制性内切核酸酶*Alu* I所识别的 AGCT，在人类基因组中约有一百多万个拷贝。

amber codon：琥珀密码子。UAG，编码终止信号。

amber mutation：琥珀突变。见 nonsense mutation。

amber suppressor：琥珀抑制子。具有能够识别琥珀密码子（UAG）的反密码子的 tRNA，能抑制琥珀突变。

amino acid：氨基酸。蛋白质的组成单元。

aminoacyl-AMP：氨酰-AMP。氨基酸通过高能酸酐键与 AMP 的磷酸基团相连接而被活化的一种结构。它是在 tRNA 负载的第一步中由氨酰 tRNA 合成酶完成。

aminoacyl-tRNA：氨酰-tRNA。通过酯化作用将自身 3′端羟基与其特定负载的氨基酸连接在一起的一种 tRNA。

aminoacyl-tRNA synthetase：氨酰-tRNA 合成酶。将 tRNA 与其特定负载的氨基酸连接在一起的酶。

a. class I：I 类。在 tRNA 的 2′-OH 端负载氨基酸的一种氨酰-tRNA 合成酶。

b. class II：II 类。在 tRNA 的 3′-OH 端负载氨基酸的一种氨酰-tRNA 合成酶。

amino tautomer：氨基互变异构体。在核酸中发现的腺嘌呤或胞嘧啶的正常的互变异构体。

amino terminus：氨基末端。一条多肽链中含有游离氨基基团的一端。蛋白质合成的起始端。

amplification：扩增。在一个单倍体基因组中，通过选择性复制使某个基因产生比正常单拷贝更多的拷贝数。

anabolic metabolism：合成代谢。由相对简单的前体物生成某种物质的过程，如色氨酸操纵子编码合成色氨酸的代谢酶。

annealing of DNA：DNA 退火。变性分离的两条 DNA 单链重新结合在一起，再次形成双链螺旋的过程。

annotated gene：被注释的基因。来自于基因组测序计划且至少部分已被鉴定的基因或基因类似序列。

antibody：抗体。能够特异识别和结合一种物质（通常为另一种蛋白质）的蛋白质。它帮助机体免疫系统识别外来入侵物质并激发免疫系统产生对入侵物进行攻击的能力。

anticodon：反密码子。tRNA 上与某个特定密码子互补配对的 3 碱基序列。

anticodon loop：反密码子环。通常位于 tRNA 分子底部含有反密码子的一个环状结构。

antigen：抗原。能被一种抗体识别和结合的物质。

antiparallel：反向平行。存在于一条 DNA 双螺旋中的两条单链的相对极性；如果从顶部到底部，一条链的方向是从 5′→3′，那么另一条链的方向则是从 3′→5′。相同的反向平行关系适用于任何双链多聚核苷酸或者寡核苷酸，也适用于 RNA 中的密码子-反密码子配对。

antirepression：抗阻遏。通过组蛋白或其他转录抑制因子来消除阻遏过程。抗阻遏是一个典型激活因子功能的一部分。

antisense RNA：反义 RNA。与 mRNA 互补的 RNA。

antiserum：抗血清。含有一种抗体或针对一类特定物质产生抗体的血清。

antiterminator：抗终止子。例如，λN 和 Q 蛋白，一种能使转录过程通读终止子而继续进行的蛋白质。

AP-1：由 1 分子的 Fos 和 1 分子的 Jun 组成的一种转录激活因子（Jun-Jun 的同源二聚体也有 AP1 的活性）。介导了对促进有丝分裂的佛波酯的响应。

AP endonuclease：AP 内切核酸酶。能在 AP 位点的 5′端切断一条 DNA 链的酶。

APE1（AP endonuclease1）：APE1（AP 内切核酸酶 1）。存在于哺乳动物细胞中一种酶类，它能利用其 3′→5′内切核酸酶的活性，校正由 β-DNA 聚合酶在碱基切补修复过程中造成的错误。

AP site：AP 位点。在 DNA 链中发生脱嘌呤或脱嘧啶的位点。

aporepressor：脱辅基阻遏蛋白。在辅阻遏物不存在时的无活性阻遏蛋白。

aptamer：适配体。具有能与另一种分子特异结合，行使核酶催化及其他活性功能的核酸或核酸区段（通常为 RNA）。

apurinic site（AP site）：无嘌呤位点（AP 位点）。DNA 链中失去嘌呤碱基的脱氧核糖。

apyrimidinic site（AP site）：无嘧啶位点（AP 位点）。DNA 链中失去了嘧啶碱基的脱氧核糖。

AraC：AraC。阿拉伯糖操纵子的负调节物。

Archaea：古菌。在生物化学和分子生物学特征上与真核生物和细菌都相似的一类原核生物。典型的古菌生活在极端热或者高盐环境中。某些古菌是能产生甲烷的严格的厌氧生物。

architectural transcription factor：结构转录因子。自身不能激活转录但帮助 DNA 弯曲，使其他激活因子结合促进转录的一类蛋白质。

ARE：见 AU-rich element。

Argonaute1（Ago1）：RITS复合体中与siRNA结合的一种Argonaute蛋白。

Argonaute2（Ago2）：哺乳动物中与复合体RISC（切割因子）活性有关的Argonaute蛋白。

Armitage：可能为RLC中的组分。已知它在RLC向RISC转换过程中起作用。

arrest（of transcription）：停滞（转录）。RNA聚合酶由于转录产物末端伸出酶外而发生的永久性暂停状态。只有当伸出的RNA末端被移除后才能重新恢复转录。

Artemis：一种能够解开在V（D）J重组过程中由Rag-1和Rag-2造成的DNA发夹结构的酶。

assembly factor：组装因子。在转录前起始复合体形成初期与DNA结合的一种转录因子，帮助其他转录因子组装到复合体上。

assembly map：组装图。体外核糖体颗粒进行自我组装过程中显示核糖体蛋白添加次序的一种示意图。

asymmetrical transcription：非对称转录。只转录多聚核苷酸双链分子指定区域中的一条链。

ATPase：ATP酶。一种水解ATP且释放能量用于其他细胞活动的酶。

attachment site：附着位点。见*att* site。

***att*B**：大肠杆菌*E. coli*基因组上的*att*位点。

attenuation：衰减。一种涉及转录提前终止的调控机制。

attenuator：衰减子。存在一个或多个结构基因上游，并且发生转录提前终止（衰减作用）的DNA区域。

***att*P**：λ噬菌体基因组上的*att*位点。

***att* site：*att*位点。**在噬菌体和宿主DNA上发生重组的位点，此位点允许噬菌体DNA作为一个原噬菌体整合到宿主基因组中。

AU-rich element（ARE）：AU-富集元件（ARE）。位于mRNA 3′端非翻译区，并且成为能使mRNA稳定性降低的miRNA的目标序列。

autonomously replicating sequence 1（ARS 1）：自主复制序列1。酵母的复制起始位点。

autoradiography：放射性自显影。使放射性样品对照相感光乳剂曝光，因而“自我照相”的技术。

autoregulation：自主调节。基因通过自身产物进行的调控。

B

BAC：细菌人工染色体。见bacterial artifical chromosome。

BAC walking：细菌人工染色体步移。对一个与起始BAC重叠最少的BAC进行测序，然后对与第二个BAC重叠最少的BAC进行测序，如此不断重复，直到一个重叠群中的所有BAC都被测序。

back mutation：回复突变。见reversion。

bacterial artifical chromosome（BAC）：细菌人工染色体（BAC）。基于大肠杆菌的F质粒构建的一种载体，能容纳长达300 000碱基的插入片段（平均插入片段为150 000碱基）。

bacteriophage：噬菌体。见phage。

baculoviruses：杆状病毒。一类含有大的、环状DNA基因组的杆状病毒。这些病毒的大部分会侵染毛虫，已被用作有效的真核基因表达载体。

barcode：生物条形码。地球上各种生物基因组中的一段相对较短的DNA序列。理论上，阅读此条形码可以快速鉴别任何已知的物种。

barrier：屏蔽。绝缘子通过阻断染色质固缩对染色体活性区域的影响来维持该区域内基因的活性，对沉默子产生的负效应。

basal level transcription：本底水平转录。II类基因在仅有通用转录因子与聚合酶II存在时获得的极低水平的转录。

base：碱基。一种环状、含氮化合物，在DNA中与脱氧核糖连接，在RNA中与核糖连接。

base excision repair：碱基切补修复。一种切补修复途径，开始先通过DNA糖基化酶去除受损碱基，接着用AP内切核酸酶切开由此生成AP位点的5′端，然后除去AP位点的糖磷脂，最后切除下游碱基，经DNA聚合酶和DNA连接酶填补缺口完成整个修复。

base pair（bp）：碱基对。一对碱基（A-T或G-C），相对地存在于DNA双链的每条单链中。

B-DNA：B型DNA。标准的沃森-克里克DNA模型，常出现在较高的相对湿度状态及溶液中。

BER：见base excision repair。

bHLH domain：碱性螺旋-环-螺旋结构域。与一个碱性的基序偶联的HLH基序。当两个bHLH蛋白通过它们的HLH基序实现二聚体化时，碱性基序在适当的位置与DNA的特异区域相互作用。bHLH蛋白像一对钳子，紧紧地抓住DNA的大沟。

bHLH-ZIP domain：碱性螺旋-环-螺旋-拉链结构域。发生二聚体化和与DNA结合的结构域。碱性区域分别与螺旋-环-螺旋结构域和亮氨酸拉链（ZIP）结构域相偶联。

bidirectional DNA replication：DNA双向复制。从一个共同的起始位点或复制起点开始同时向两个不同方向进行复制，需要两个活性复制叉。

bioinformatics：生物信息学。建立并运行生物数据

库的领域。在基因组学层面，它意味着管理大量的序列数据以及提供有用的数据存取和解释。

biolistics transformation (or transfection)：微弹轰击转化（或转染）。一种用金属微球包裹 DNA 并将其射入细胞的方法。

bit score (S)：二进制位分值。度量被查询的序列与从 BLAST 搜索中获得的数据库序列之间匹配数目的方法。

BLAST：局部相似性基本搜索工具（basic local alignment search tool）。一种搜索 DNA 序列或蛋白质序列数据库，并成排显示被查询的序列与数据库中的序列间同源性的程序。

branch migration：分支迁移。在重组过程中 Holliday 联接体的分支移端。

branchpoint-bridging protein (BBP)：分支点桥联蛋白 (BBP)。在内含子 5′端结合 U1 snRNP，3′端结合 Mud2p，并对剪接过程是必不可少的一种蛋白质。

bridge helix：桥螺旋。一种紧邻细菌 RNA 聚合酶活性中心，其弯曲可以促进转录转位的 α 螺旋。

BRG1：SWI/SNF 上具有 ATP 酶及染色质重建的催化活性亚基。

BRG1-associated factor (BAF)：BRG1 关联因子 (BAF)。与 BRG1 共同形成 SWI/SNF 的 9～12 条多肽。

bromodomain：溴域。一种与其他蛋白质（如组蛋白）上的乙酰化赖氨酸残基特异性结合的蛋白质结构域。

bZIP domain：碱性亮氨酸拉链结构域。与一个碱性基序偶联的亮氨酸拉链基序。当两个碱性亮氨酸拉链蛋白通过它们的亮氨酸拉链二聚体化时，碱性基序处于与 DNA 的特定区域相互作用的位置。碱性亮氨酸拉链蛋白像一对钳子紧紧地抓住 DNA 的大沟。

C

C value：C 值。在一个特定物种的单倍体基因组中的 DNA 的含量，常以 1pg 为计量单位。

C value paradox：C 值悖论。指一个特定物种的 C 值与其遗传复杂程度不一致的现象。

cAMP response element (CRE)：cAMP 响应元件。对 cAMP 发生响应的增强子。

cap：帽子（成熟 mRNA 5′端的结构）。一个甲基化的鸟嘌呤核苷通过一个三磷酸连接到 mRNA、hnRNA 或 snRNA 的 5′端上。

cap 0：帽子 0。在任何 2 位均无甲基化的帽子，只存在于某些病毒 mRNA 中。

cap 1：帽子 1。一种典型的帽子，在倒数第二个核苷酸的 2 位羟基上有一个甲基。

cap 2：帽子 2。在少数 mRNA 中发现的一种在前两个核苷酸的 2 位羟基上都有甲基的帽子。

CAP (catabolite activator protein)：代谢物激活因子蛋白。一种与 cAMP 一起激活受代谢物阻遏的操纵子的蛋白质，也称之为 CRP。

cap-binding protein (CBP)：见 eIF4F。

carboxyl-terminal domain (CTD, of Rpb1)：羧基末端结构域（Rpb1 上的 CTD）。RNA 聚合酶 II 最大亚基的羧基末端区域。由多个富含丝氨酸和苏氨酸的七聚体重复组成。

carboxyl terminus：羧基末端。多肽链具有自由羧基的末端。

CARM1：见 coactivator-associated arginine me-thyltransferase。

catabolic metabolism：分解代谢。将物质分解为简单成分的过程，如乳糖操纵子编码的分解代谢酶将乳糖分解成半乳糖和葡萄糖。

catabolite activator protein：代谢物激活因子蛋白。见 CAP。

catabolite repression：代谢物阻遏作用。葡萄糖或代谢物或葡萄糖的降解产物对一个基因或操纵子的阻遏作用。

catalytic center：催化中心。酶的活性位点即催化反应的发生地。

catenane：环连体。由两个或多个环连接成链状的结构。

CBP：见 CREB-binding protein。

CCAAT-binding transcription factor (CTF)：与 CCAAT 结合的转录因子 (CTF)。结合在 CCAAT 框上的一种转录因子。

CCAAT box：CCAAT 框。具有 CCAAT 序列，在许多被 RNA 聚合酶 II 识别的真核生物启动子中发现的一种上游基序。

Cdc13p：结合在单链端粒末端并募集 Stn1p 因子（而 Stn1p 招募 Ten1p 结合到端粒末端）的一种酵母蛋白。3 种蛋白质共同保护端粒末端免遭降解以及 DNA 修复酶的作用。

cDNA：互补 DNA (complementary DNA)。经反转录合成的一种 RNA 的 DNA 拷贝。

cDNA library：cDNA 文库。包含在特定时间内的一类细胞中所有可能的 mRNA 的一套 DNA 克隆。

centimorgan (cM)：厘摩 (cM)。在两个标记之间发生 1%重组频率的遗传距离。

centromere：着丝粒。染色体上的固缩区域，在细胞分裂时纺锤丝的附着处。

CF I and CFII：见 cleavage factor I and II。

chaperone protein：**分子伴侣蛋白**。见 chaperones。

chaperones：**分子伴侣**。与未折叠的蛋白质结合并帮助其进行正确折叠的蛋白质。

charging：**负载**。将特定的氨基酸与 tRNA 偶联的事件。

Charon phage：卡隆噬菌体。基于 λ 噬菌体构建的一套克隆载体。

Chi site：**Chi 位点**。具有 5′-GCTGGTGG-3′保守序列的大肠杆菌 DNA 的位点。在同源重组过程中 RecBCD 切割 chi 位点的 3′端。

Chi structure：见 Holliday junction。

ChIP：见 chromatin immunoprecipitation。

chloramphenicol：**氯霉素**。通过抑制 SOS 核糖体催化的肽基转移酶反应杀死细菌的一种抗生素。

chloramphenicol acetyl transferase (CAT)：**氯霉素乙酰转移酶 (CAT)**。可将乙酰基团转移到氯霉素上的一种酶。其对应的细菌基因常被用作真核生物转录和翻译研究中的报告基因。

chromatid：**染色单体**。在细胞分裂中染色体复制产生的一个拷贝。

chromatin：**染色质**。由 DNA 和染色体蛋白组成的染色体物质。

chromatin immunoprecipitation (ChIP)：**染色质免疫沉淀技术 (ChIP)**。一种通过免疫沉淀来纯化含有目的蛋白的染色质的方法，该方法是用一种含有针对该目的蛋白的，或者针对吸附有该目的蛋白的抗原决定簇标签的抗体来免疫沉淀这种待纯化的染色质。

chromatin remodeling：**染色质重建**。依赖 ATP 的核小体结构重建过程可以移动核小体，或这些核小体也能被其他蛋白质所移动。

chromatography：**层析**。基于流动相与固定相的相对亲和力来分离不同分子的一套技术。在离子交换层析中，带电的树脂是固定相，提高了离子强度的缓冲液是流动相。

chromodomain：**染色域**。在与异染色质形成有关的蛋白质上发现的保守区域。该区域可能与甲基化的组蛋白相结合。

chromogenic substrate：**显色底物**。当被酶作用后能生成有色产物的一种底物。

chromosome：**染色体**。由大量 DNA 和蛋白质构成的物理结构，包含有机体的全部基因。

chromosome theory of inheritance：**遗传的染色体理论**。基因存在于染色体上的学说。

cI：编码 λ 阻遏蛋白的基因。

CI：***cI* 基因的产物**。见 λ repressor。

***cis*-acting**：**顺式作用**。一种描述增强子、启动子或操纵子等遗传因子的术语，为了影响某个靶基因的活性它们必须位于相同的染色体上。

***cis*-dominant**：**顺式显性**。位于同一 DNA 片段上的基因的显性效应。例如，在含有 *lac* 操纵子部分二倍体的大肠杆菌中，一个拷贝中的操纵基因的组成型突变对另一拷贝的操纵子是显性。对其他操纵子不具有该效应。因为此操纵基因只控制着与它相邻接的操纵子，而不能控制与之不相连的操纵子。

***cis*-splicing**：**顺式剪接**。将来源于同一个前体 RNA 分子中的外显子连接的一种常规剪接方式。

cistron：**顺反子**。通过顺-反测验所定义的一个遗传因子。它通常与基因是同义词。

clamp loader：**钳状装载器**。DNA 聚合酶 III 全酶的 γ 复合体部分，它帮助 β 钳结合到 DNA 上。

clamp module：**钳状模块**。RNA 聚合酶的一部分，其张开允许 DNA 模板进入，闭合则将聚合酶夹在模板上。

class II promoter：**II 类启动子**。识别 RNA 聚合酶 II 的启动子。

clastogen：**断裂剂**。可引起 DNA 链断裂的试剂。

cleavage factor I and II (CF I and CF II)：**剪切因子 I 和 II (CF I and CF II)**。RNA 的结合蛋白，它们对前体 mRNA 在多聚腺苷酸化位点处进行剪切很重要。

cleavage poly (A) specificity factor (CPSF)：**多聚腺苷酸剪切特异因子 (CPSF)**。一种能够识别前体 mRNA 中多聚腺苷酸化信号 AAUAAA 的蛋白质，并激活剪切和多聚腺苷酸化。

cleavage stimulation factor (CstF)：**剪切刺激因子 (CstF)**。一种能够识别前体 mRNA 多聚腺苷酸化信号中富含 GU 部分的蛋白质，并促进剪切。

CLIM (cofactor of LIM)：**CLIM (LIM 的辅因子)**。LIM-HD 的辅因子。在被 RLIM 泛素化后该辅因子被蛋白质降解。

clone-by-clone sequencing：**克隆步移测序**。对大基因组测序的一种系统方法。首先对整个基因组作图，然后对基因组已知区域的克隆进行测序。

clone：**克隆**。通过无性繁殖过程形成的很多个体群，以保证它们在遗传上与最初的个体（祖先）是相同的。同样，菌落或细胞群或病毒群在遗传上也都是一致的。

closed promoter complex：**封闭型启动子复合体**。由 RNA 聚合酶和原核启动子之间相对松散结合形成的复合物。“封闭型”意指 DNA 双链仍保持完整，双链没有“打开”或碱基对没有分离。

coactivator-associated arginine methyltransferase (CARM1)：辅激活因子关联的精氨酸甲基转移酶。一种在启动子附近使蛋白质甲基化从而激活转录的真核生物蛋白质。

coactivator：辅激活因子。本身没有转录激活能力，但能帮助其他蛋白质激活转录的因子。

codon：密码子。mRNA 中连续排列的三碱基序列，能使一个特异的氨基酸插入到蛋白质中或引起翻译的终止。

coiled coil：卷曲螺旋。两个 α 螺旋相互缠绕的一种蛋白质基序。当两个 α 螺旋分别在不同蛋白质上时，卷曲螺旋的形成会导致二聚体化。

cointegrate：共联体。转座子在转位过程中产生的中间体，如 Tn3 转座子从一个复制子转座到另一个复制子。当转座子发生复制，共合体就包含了通过两个转座子拷贝连接在一起的两个复制子。

colE1：在大肠杆菌 *E. coli* 的某菌株中发现的一种质粒，它编码一种称之为大肠杆菌素的细菌毒素。colE1 的 DNA 复制是单方向性的。

colony hybridization：菌落杂交。一种筛选含有目的基因的细菌克隆的方法。用目的基因标记的探针对大量的克隆的 DNA 进行杂交 met 测试。

combinatorial code：组合密码。对多个增强子和相应激活因子在其结合的启动子上的作用的一种描述。结合不同的激活因子将对启动子的活性产生不同的影响，因为增强子能感受每一个激活因子的浓度并将来自所有激活因子的信号整合在一起。

commitment complex (CC)：定向复合体 (CC)。一种至少含有细胞核前体 mRNA 和 U1 snRNP 的复合物，负责把 U1 snRNP 结合的内含子剪切掉。

complementary polynucleotide strand：互补多聚核苷酸链。两条 DNA 链或具有互补序列的 RNA 链，即无论哪条链含有腺嘌呤则另一条链就含有胸腺嘧啶，同样，一条链含有鸟嘌呤另一条链就含有胞嘧啶。

complex B：复合体 B。RLC 的前体，仅含有 Dicer 和 R2D2 以及双链 siRNA。

composite transposon：复合转座子。一种由两个不同部分组成的细菌转座子：含有 IS 或类 IS 元件的两臂，以及由转位基因和一个或多个抗生素抗性基因组成的中心区域。

concatemer：多联体。具有多个基因组长度的 DNA。

conditional lethal：条件性致死。在某种环境下为致死突变，但在其他环境下却不致死（如温度敏感型突变）。

consensus sequence：共有序列。几个共有的相似序列，如 *E. coli* 启动子 −10 框的保守序列是 TATAAT，即如果你检测许多这样的序列，T 最有可能出现在第个位置，A 在第二个位置等（如此类推）。

conservative replication：全保留式复制。两条亲代链仍然连在一起的 DNA（或 RNA）复制，所产生的子链的双链都是新合成的。

conservative transposition：保守式转座。当转座子离开原来的位置并且移动到新位点时，转座子 DNA 的两条链仍被保留完整的一种“剪贴式”转座。

constant region：恒定区。抗体上的一段区域，或多或少与其他抗体相似。

constitutive：组成型的。总是处于开启状态。

constitutive mutant：组成型突变体。一个有机体中的组成型突变能引起某个基因在任何时间都被表达，而不被正常途径所调控。

contig：重叠群。一组含有毗邻或重叠序列的 DNA 克隆。

***copia*：**在果蝇 *Drosophila* 细胞中发现的一种能转座的元件。

core element：核心元件。一种被 RNA 聚合酶 I 所识别的真核启动子元件。包括围绕在转录起始位点两侧的碱基。

core histone：核心组蛋白。除 H1 之外的所有核小体组蛋白。被 DNA 缠绕的核小体内部的组蛋白。

core polymerase：核心聚合酶。见 RNA polymerase core。

corepressor：辅阻遏物。与阻遏物蛋白结合形成有阻遏活性的物质（如色氨酸是色氨酸操纵子 *trp* 的辅阻遏物）。与其他蛋白质结合对基因转录起抑制作用的蛋白质，如组氨酸去乙酰化酶能作为辅阻遏物。

core promoter (class II)：核心启动子 (II 类)。在一个特定的启动子中靠近转录起始位点的任何启动子元件。最多包括 4 种元件：一个 TFIIB 识别元件、一个 TATA 盒、一个起始位点和一个下游元件。

core promoter element：核心启动子元件。一个启动子包含的最少元件（如细菌启动子中的−10 框和−35 框）。

cos：线性 λ 噬菌体 DNA 的黏性末端。

cosmid：cos 质粒。一种为了克隆 DNA 大片段而设计的载体。cos 质粒含有 λ 噬菌体的 cos 位点，所以能被包装进入 λ 噬菌体的头部，它又含有质粒的复制起点，所以能像质粒那样复制。

CoTC element co-transcriptional cleavage (CoTC)：CoTC 元件共转录剪切 (CoTC)。对多聚腺苷酸化位点下游正在产生的转录物进行剪切，是转录终

止过程的一部分。

count per minute (cpm)：每分钟的记数。液体闪烁计数器在每分钟内检测到的光闪烁的平均值。

CPEB：胞质多聚腺苷酸化元件（CPE）结合蛋白。

CpG island：CpG 岛。含有许多非甲基化 CpG 序列的 DNA 区域。通常与活性基因相连接。

CpG sequences：CpG 序列。哺乳动物 DNA 中甲基化标靶（在 C 的 5 位）的基序。

CpG suppression：CpG 抑制。由于 C 的甲基化及随后 T 的去氨基化因素，在漫长的进化历程中最终导致 CpG 序列从基因组中丢失。

CPSF：见 cleavage and poly（A）specificity factor。

CRE：见 cAMP response element。

CREB：见 CRE-binding protein。

CREB-binding protein（CBP）：CREB 结合蛋白（CBP）。一种辅激活因子，在 CRE 内结合被磷酸化的 CREB，然后结合一个或多个通用转录因子，从而刺激前起始复合物的装配。

CRE-binding protein（CREB）：CRE 结合蛋白（CREB）。磷酸化从而被 cAMP-刺激的蛋白激酶 A 所活化的激活因子。结合 CRE 并与 CBP 共同激活相关基因的转录。

Cro：λ 噬菌体 *cro* 基因的产物。一种优先与O_R3结合并关闭 λ 阻遏蛋白基因（*cI*）的抑制因子。

cross-linking：交联。探查两种物质之间相互作用的一种技术（如一种蛋白质和一种 DNA）。当两种物质形成复合体时就会产生化学交联，然后对产物的交联特性进行检测。

crossing over：交换。在重组过程中 DNA 之间产生的物理交换。

cross talk：交互作用。不同信号转导通路成员之间的相互作用。

crown gall：冠瘿瘤。在植物中由细菌感染引起的肿瘤性的增生。

CRP：环化-磷酸腺苷酸受体蛋白。见 CAP。

CRSP：与 Sp1 协作共同激活转录的辅激活因子。

cryptogene：隐秘基因。编码一种需要进行转录后编辑的 RNA 的基因。

CstF：见 cleavage stimulation factor。

CTD：见 carboxyl-terminal domain。

cyanobacteria（blue-green algae）：蓝细菌（蓝绿藻）。光合细菌。现代蓝绿藻的祖先被认为能侵入真核细胞并进化成叶绿体。

cyclic-AMP（cAMP）：环化-磷酸腺苷酸（cAMP）。一种通过环磷酸二酯键连接 3′端和 5′端碳原子的腺嘌呤核苷酸。涉及原核生物和真核生物的调控机制。

cytidine：胞嘧啶核苷酸。含有胞嘧啶碱基的一种核苷酸。

cytidine deaminase acting on RNA（CDAR）：作用于 RNA 的胞嘧啶核苷酸脱氨酶（CDAR）。一种将 RNA 中的某些胞嘧啶核苷脱去氨基并转化为尿嘧啶核苷的 RNA 编辑酶。

cytoplasmic polyadenylation element（CPE）：细胞质多聚腺苷酸化元件（CPE）。mRNA 3′端 UTR 中的一段序列（保守序列为 UUUUUAU），在细胞质多聚腺苷酸化中很重要。

cytosine（C）：胞嘧啶（C）。在 DNA 中与鸟嘌呤核苷配对的嘧啶碱基。

cytotoxic：细胞毒性。具杀死细胞的能力。

D

DAI：见 double-stranded RNA-activated inhibitor of protein synthesis。

***dam* methylase：*dam* 甲基化酶。**脱氧腺嘌呤核苷甲基化酶，在大肠杆菌细胞中能将甲基基团转移到 GATC 序列的 A 上。错配修复系统能扫描甲基化的 GATC 序列，并确定哪条链是新合成的，并且是未被甲基化的。

daughter strand gap：姐妹链缺口。DNA 复制装置在跳过非编码的碱基或一个嘧啶二聚体之后留下的缺口。

deadenylation：去腺苷酸化。从细胞质的poly（A)中切除 AMP 残基。

DEAD protein：DEAD 蛋白。含有 Asp-Glu-Ala-Asp 序列的蛋白质家族成员，具有 RNA 解旋酶活性。

deamination of DNA：DNA 的脱氨基化。从 DNA 的胞嘧啶或腺嘌呤中切除氨基（NH_2），并被羰基（C—O）所代替。这将使胞嘧啶转化为尿嘧啶以及腺嘌呤转化为次黄嘌呤。

decatenation：去连环。解除连环体中各环之间连接的过程。

decoding：解码。核糖体通过密码子与反密码子间的相互作用，引导结合一个正确的氨酰 tRNA。

defective virus：缺陷型病毒。没有辅助病毒就不能复制的一种病毒。

degenerate code：简并密码。地球上所有生物广泛使用的遗传密码，此体系中多个密码子能对应同一种氨基酸。

deletion：缺失。包括丢失一个或多个碱基对的突变。

denaturation（DNA）：变性（DNA）。DNA 双链的分离。

denaturation（protein）：变性（蛋白）。在没有打断

任何共价键的情况下蛋白质三维结构的破坏。

densitometer：显像密度计。一种测量透明薄膜上黑点的仪器（如放射性照相）。

deoxyribose：脱氧核糖。DNA 中的糖。

Dicer：RNase III 家族的成员，在 RNAi 的过程中将引发 RNA 切短成大约 21bp 的片段，也是降解靶 mRNA 的 RISC 复合体的一部分。

dideoxyribonucleotide：双脱氧核糖核苷酸。一种核苷酸，在它的 2′和 3′的位置上都脱去了氧，在 DNA 测序中用来终止 DNA 链的延伸。

dihydrouracil loop：二氢尿嘧啶环。见 D loop。

dimer（protein）：二聚体（蛋白）。两条多肽链的形成的复合物。这些多肽可以是相同的（同源二聚体），或者不同（异源二聚体）。

dimerization domain：二聚体化结构域。一个蛋白质与另一个蛋白质相互作用形成二聚体（或者更高级的多聚体）的部位。

dimethyl sulfate（DMS）：硫酸二甲酯（DMS）。可以使 DNA 发生甲基化的一种化学试剂。在甲基化位点上 DNA 能被化学剪切。

diploid：二倍体。人类受精卵和其他细胞（除配子之外）的染色体数目。用 $2n$ 代表双倍体的符号。

directional cloning：定向克隆。将外源 DNA 插入到一个载体的两个不同限制酶酶切位点之间，能预先确定这种插入的方向。

disintegrations per minute（dpm）：每分钟衰变数（dpm）。一个样品平均每分钟产生的放射性发射数。

dispersive replication：散布式复制。是一种假设机制，复制之后 DNA 变成片段，使新旧 DNA 共存于同一条链中。

distributive：分配性的。是相对于“持续性”而言的，如果没有与底物或模板反复地分离和再结合则不能完成某项任务。

D loop：D 环。当一条 DNA 或 RNA 链末端“入侵”到一条双螺旋中，并与其中一条链的碱基配对，从而迫使另一条链凸出（loopout）时形成的环。

DMS footprinting：DMS 足迹法。一种与 DNase 足迹法相似的技术，它是用 DMS 使 DNA 甲基化后再进行化学剪切的方法，而不是利用 DNase 剪切 。

DNA（deoxyribonucleic acid）：DNA（脱氧核糖核苷酸）。一种通过磷酸二酯键将脱氧核糖核苷酸连接在一起的聚合物。大多数基因是由此物质构成的。

DnaA：在形成的 *E. coli* 引发体中，与 *ori*C 结合的第一个蛋白质。

***dnaA* box：*dnaA* 框。**在形成的 *E. coli* 引发体中，位于与 DnaA 结合的 *ori*C 内部的一段 9 碱基重复区。

DnaB：*E. coli* 引发体的一个关键成分。通过促进引发酶的结合来帮助引发体的组装。也具有 DNA 解旋酶活性，在引物合成之前解开母链 DNA。

DNA-binding domain：DNA 结合的结构域。蛋白质中与 DNA 发生结合的部分，与 DNA 的靶位点形成特异性的连接。

DNA fingerprint：DNA 指纹。用 DNA 的高变化区域来鉴定特定的个体。

DnaG：*E. coli* 的引发酶。

DNA glycosylase：DNA 糖基化酶。一种能打断受损碱基与其糖基间的糖苷键的酶。

DNA gyrase：DNA 促旋酶。将负超螺旋导入（泵进）DNA 的一种拓扑异构酶。在复制过程中通过对 *E. coli* DNA 的解链造成正超螺旋的松弛。

DNA ligase：DNA 连接酶。使两条 DNA 双链首尾相连接的酶。

DNA melting：DNA 的熔解。见 denaturation（DNA）。

DNA microarray：DNA 微阵列。含有许多 DNA 或寡聚核苷酸的微小点阵的一种芯片。用于斑点印迹可同时检测许多基因的表达。

DNA microchip：DNA 微芯片。见 DNA microarray。

DNA photolyase：DNA 光复活酶。一种通过打断嘧啶二聚体来催化光复活作用的酶。

DNA-PK（DNA protein kinase）：DNA-PK（DNA 蛋白激酶）。在真核生物中修复双链断裂的关键酶。

DNA-PK$_{cs}$：DNA-PK 的催化亚基。

DNA polymerase：DNA 聚合酶。合成 DNA 的酶，它通过与模板 DNA 链中的核苷酸序列的互补来指导磷酸脱氧核糖核苷（dNMP）间的有序地连接。

DNA polymerase η：DNA 聚合酶 η。一种特异的真核跨损伤合成聚合酶，复制过程中负责在嘧啶二聚体模板的对面插入两个 dAMP。

DNA polymerase θ：DNA 聚合酶 θ。一种特异的真核 DNA 聚合酶，参与跨损伤合成。

DNA polymerase I（pol I）：DNA 聚合酶 I（pol I）。*E. coli* 3 种不同 DNA 合成酶之一，主要用于 DNA 修复。

DNA polymerase II（pol II）：DNA 聚合酶 II（pol II）。*E. coli* 的另一种 DNA 聚合酶。

DNA polymerase III holoenzyme：DNA 聚合酶 III 全酶。*E. coli* 复制体中的酶，在 DNA 复制过程中真正负责合成 DNA 的酶。

DNA polymerase ζ：DNA 聚合酶 ζ。在人类细胞中参

与跨损伤合成的一种特异性 DNA 聚合酶。

DNA polymerase V：DNA 聚合酶 V。见 $UmuD'_2C$。

DNA protein kinase：DNA 蛋白激酶。见 DNA-PK。

DNase：脱氧核糖核酸酶。一种降解 DNA 的酶。

DNase footprinting：DNase 足迹法。通过观察受蛋白质保护而不被 DNase 降解的 DNA 区域，来检测此蛋白质在 DNA 上的结合位点的方法。

DNase-hypersensitive site：DNase 超敏感位点。受 DNase I 的攻击程度较松散染色质敏感 100 倍以上的染色质区域。这些位点通常位于活性或具有潜在活性基因的 5′侧区。

DNase-sensitive site：DNase 敏感位点。对DNase I的敏感程度较松散染色质高 10 之多倍的染色质区域。所有活跃基因都趋于对 DNase 敏感。

DNA typing：DNA 分型。运用分子技术特别是 Southern 印迹技术来鉴定个体的特征。

dominant：显性。一个等位基因或它的表型特征与具有一个隐性等位基因的杂合体所表现出的性状相同。例如，*A* 相对于 *a* 是显性，因为 *AA* 和 *Aa* 的表型都是一样的。

dominant-negative mutation：显性负突变。一种基因突变，它产生的蛋白质不仅失去活性，而且能通过形成混合多聚体方式破坏同一细胞中野生型蛋白质的活性。

domain（protein）：结构域（蛋白）。蛋白质中的一个独立折叠部分。

double helix：双螺旋。染色体中两条互补 DNA 链所呈现的形状。

double-stranded RNA-activated inhibitor of protein synthesis（DAI）：蛋白质合成中双链 RNA 激活的抑制因子（DAI）。对干扰素和双链 RNA 做出应答的一种蛋白激酶，通过使 eIF-2α 磷酸化并增强其对 eIF-2B 的结合从而阻断翻译的起始。这将阻止感染细胞中病毒蛋白质的合成。

down mutation：下降突变。通常发生在启动子中的一种突变，导致某个基因的表达下降。

downstream destabilizing element：下游去稳定元件。在 mRNA 成熟过程的剪接步骤中，结合到外显子-外显子连接处的一系列蛋白质。细胞利用这些蛋白质作为参照来确定一个无义密码子是真正的终止密码还是提前终止密码。

downstream promoter element（DPE）：下游启动子元件。转录起始位点下游的一段 DNA 区域，对一些 II 类启动子的转录效率十分重要。

***Drosophila melanogaster*：**黑腹果蝇，被遗传学家广泛利用的一种果蝇。

***Ds*：**在玉米中发现的一种缺陷型转座因子，依赖于 *Ac* 因子才能进行转座。

DSB：DNA 中的双链断裂。在减数分裂过程中对启动重组事件是必需的。

DskA：一种细菌蛋白，在饥饿响应中，与预警子 ppGpp 联合可减缓 rRNA 的合成。

dsRNA：双链 RNA。

dUTPase：脱氧三磷酸尿嘧啶核苷酸水解酶。降解 dUTP 从而阻止其掺入 DNA 的酶类。

E

editing：编辑。发生在氨酰-tRNA 合成酶编辑位点上的一个事件，用以终止被合成酶错误激活的非特定的氨酰-AMP（或氨酰-tRNA）掺入的过程。也被称为校正。

editing site：编辑位点。氨酰-tRNA 合成酶中的一个位点，用以查验氨酰腺苷酸化（有时是氨酰-tRNA）并对那些带有太小氨基酸的氨酰-tRNA 进行水解。

EF-2：延伸因子 2。延伸因子 G 的真核生物同源物。

EF-G：延伸因子 G。与 GTP 一起负责在原核生物的翻译过程中转位的延伸因子。

EF-Ts：延伸因子 Ts。将延伸因子 Tu 上的 GDP 交换为 GTP 的交换因子。

EF-Tu：延伸因子 Tu。与 GTP 一起携带氨酰-tRNA（除甲酰甲硫氨酰-$tRNA_i^{Met}$ 外）进入到核糖体的 A 位点的原核生物翻译延伸因子。

EGF：见 epidermal growth factor。

eIF1：真核生物起始因子 1。通过激发扫描过程找到正确起始密码子的真核生物起始因子。

eIF1A：真核生物起始因子 1A。与 eIF1 协同作用使 40S 核糖体亚基扫描起始密码子。

eIF2：真核生物起始因子 2。负责将甲硫氨酰-$tRNA_i^{Met}$ 结合到核糖体 40S 亚基上的真核生物起始因子。

eIF2α：真核生物起始因子 2α。eIF2 的亚基之一。它被磷酸化后可以抑制翻译起始。

eIF2B：真核生物起始因子 2B。将 eIF2 上的 GTP 转换为 GDP 的真核交换因子。

eIF3：真核生物起始因子 3。结合在 40S 核糖体亚基上的真核生物起始因子，能阻止它们与 60S 核糖体亚基过早地再结合。

eIF4A：真核生物起始因子 4A。eIF4F 的亚基之一。是具有 RNA 解旋酶活性的 DEAD 家族的 RNA 结合蛋白。与 eIF4B 连接后可结合到 mRNA 的前导区，并在扫描核糖体亚基之前除去发夹结构。

eIF4B：真核生物起始因子 4B。在翻译起始中帮助 eIF4A 结合到 mRNA 上的 RNA 结合蛋白。

eIF4E：真核生物起始因子 4E。eIF4F 中的帽子结合组分。

eIF4F：真核生物起始因子 4F。真核生物中参与转录过程的帽子结合复合体。

eIF4G：真核生物起始因子 4G。eIF4F 的亚基之一。是连接两个不同蛋白质的连接因子：与已结合在帽子上的 eIF4E 结合；同时也与已结合在 40S 核糖体颗粒上的 eIF3 结合。通过这种方式，使 40S 小亚基与 mRNA 的 5′端结合并开始扫描。也与结合 poly（A）的 PABI 结合。

eIF5：真核生物起始因子 5。促使 40S 起始复合体和 60S 核糖体亚基结合的真核生物起始因子。

eIF5B：真核生物起始因子 5B。真核生物中与原核生物 IF2 同源的因子。帮助 eIF5 募集 60S 核糖体亚基到起始复合体。需要 GTP 水解才能从核糖体被释放。

eIF6：真核生物起始因子 6。与 eIF3 活性相似的真核生物起始因子。

electron-density map：电子密度图。对单个分子或复合分子中电子密度的三维表示法。

electrophile：亲电子试剂。在其他分子中寻找负电荷中心并在那里进行攻击的一种分子。

eletrophoresis：电泳。将电压施加到带电分子上并引导其移动的过程。利用这一技术可以分离 DNA、RNA 或蛋白质片段。

electrophoretic mobility shift assay（EMSA）：电泳迁移率变化测验。见 gel mobility shift assay。

electroporation：电穿孔法。用一股强电流将 DNA 导入细胞内。

Elk-1：一种激活因子，也是丝氨酸/苏氨酸激酶信号转导的标靶分子。

elongation factor：延伸因子。与氨酰-tRNA 结合并在翻译延伸阶段的转位步骤中必需的蛋白质。

encode：编码。含有合成 RNA 或多肽的信息。一个基因可以编码一个 RNA 或一条多肽。

end-filling：末端填充。利用三磷酸脱氧核苷和 DNA 聚合酶填补双链 DNA 3′凹陷端。这种技术常常被用于标记 DNA 链的 3′端。

endonuclease：内切核酸酶。在多聚核苷酸链内部进行切割的酶。

endoplasmic reticulum（ER）：内质网。字面意思为“细胞内的网状结构”，即在细胞内合成并输出蛋白质的膜网状结构。

endospores：内生孢子。在细胞内形成的休眠孢子，像枯草杆菌。

enhanceosome：增强小体。由增强子与激活因子偶联形成的复合体。

enhancer：增强子。促进一个或多个基因转录的 DNA 元件。增强子一般存在于其调控基因的上游，但它们也可以颠倒或在间距几百甚至上千对碱基之外对基因起调控作用。

enhancer-binding protein：增强子结合蛋白。见 activator。

enhancer-blocking：增强子屏障。对增强子具有负效应的绝缘子。

enzyme：酶。通常是一种蛋白质但有时是 RNA 的分子，可以催化或加速并引导一个生化反应。

E1，E2 and E3 snoRNA：E1、E2 和 E3 小核仁 RNA。见 small nucleolar RNA。

epidermal growth factor（EGF）：表皮生长因子（EGF）。结合到跨膜受体上给细胞传递分裂信号的一类蛋白质。

epigenetic：表观遗传。没有影响 DNA 碱基序列改变的表型遗传。

epitope tagging：抗原决定簇标签法。用遗传学方法将一小段氨基酸残基（抗原决定簇标签法）融合到目的蛋白上。这可使蛋白质通过抗原决定簇标签法识别抗体进行免疫沉淀进而被纯化。

ERCC1：人类切补修复交叉互补基因 1。与 DNA 修复基因 XPF 一起，在单核苷酸切补修复中切除损伤 DNA 的 5′端。

eRF1：真核生物释放因子 1。这种因子能识别 3 种终止密码子并从核糖体中释放出合成完毕的多肽。

eRF3：真核生物释放因子 3。具有依赖核糖体的 GTP 酶活性，与 eRF1 协同从核糖体中释放合成完毕的多肽。

ERK（extracellular signal-regulated kinase）：细胞外信号调节激酶。一个由 MEK 激活的信号转导丝氨酸/苏氨酸蛋白激酶，然后在细胞核中激活 Elk-1等激活因子。

error-prone bypass：易错旁路。细胞用嘧啶二聚体或非编码碱基复制 DNA 的一种机制。将在 DNA 损伤修复过程中易于引发突变的 DNA 聚合酶（易错 DNA 聚合酶）募集到损伤处并随机插入核苷酸。

***Escherichia coli*（*E. coli*）：大肠杆菌。**一种肠内细菌，是研究细菌分子生物学的最佳材料。

E site（ribosomal）：出口位点（核糖体）。卸载了酰基的 tRNA 在离开核糖体时所结合的出口位点。

E site（RNA polymerase）：出口位点（RNA 聚合酶）。后续进入的核苷酸在转入（包括旋转）A 位点之前所占据的位点。

essential gene set：必需基因族。一组基因，其丢失会危及一个有机体的生命。

EST：序列表达标签。见 expressed sequence tag。

euchromatin：常染色质。处于伸展状态的染色质，RNA 聚合酶容易进入，至少具有潜在的活性，染色浅或正常，被认为含有大多数基因。

eukaryote：真核生物。细胞中有细胞核的一类有机体。

evolutionarily conserved region（ECR）：进化保守区域（ECR）。在许多生物体的 DNA 中所具有的序列，可能位于外显子上。

excinuclease：切除核酸酶。在人的单核苷酸切补修复中参与切除发生损伤的寡核苷酸的一类内切核酸酶。

excision repair：切补修复。对损伤 DNA 的修复，涉及去除损伤 DNA 并以正常 DNA 替代的修复。

exon：外显子。最终代表基因或成熟转录产物的一个基因区域。这个词可同时用于 DNA 及其 RNA 产物。

exon definition：外显子界定。剪接因子识别外显子末端的剪接方案。

exonic splicing enhancer（ESE）：外显子剪接增强子。外显子中一段能促进剪接的区域。

exonic splicing silencer（ESS）：外显子剪接沉默子。外显子中一段抑制剪接的区域。

exon trapping：外显子捕获法。通过插入任意 DNA 片段到载体中进行表达来克隆外显子的一种方法，只有完整的外显子才能表达。

exonuclease：外切核酸酶。从末端向内降解多聚核苷酸的一类酶。

exosome：外切酶体。降解 RNA 的一种蛋白质复合体，在细胞核和细胞质中发现了不同的外切酶体。

expect value（*E* value）：期望值（*E* 值）。预期随机看见产生相应比特值的匹配数。期望值越低，匹配越好。

expressed sequence tag（EST）：表达序列标签。通过 RT-PCR 扩增细胞内的 mRNA 而产生的一种 STS（序列标签位点）。

expression vector：表达载体。使被克隆基因表达的一种克隆载体。

F

F plasmid：F 质粒。使细菌细胞间产生接合的一类大肠杆菌质粒。

F′ plasmid：F′质粒。含有一段宿主 DNA 的 F 质粒。

F_1：子一代。一个或多个不同基因型的亲本间杂交的后代。

F_2：子二代。两个 F_1 个体杂交或一个 F_1 个体自交的后代。

FACT（facilitates chromatin transcription）：促染色质转录的蛋白质因子。体外通过核小体促进转录的一类蛋白质。能与组蛋白 H2A 和 H2B 强烈互作，去除这两个核心组蛋白可能使核小体趋于不稳定。

far Western blot：与 Western blot 相似的印迹，但以标记的蛋白质（而不是抗体）作探针，此蛋白质可能结合到印迹的蛋白质上。

ferritin：铁蛋白。细胞内的一类铁存储蛋白。

fingerprint（protein）：指纹（蛋白）。当一个蛋白质被一个酶（如胰蛋白酶）切成片段（肽段）后形成的特异性肽段，然后通过层析可分离这些肽段。

FISH（fluorescence *in situ* hybridation）：荧光原位杂交。用一个荧光探针与整条染色体杂交来确定一个基因或其他 DNA 序列在染色体上所在位置的方法。

flap-tip helix：瓣顶螺旋。大肠杆菌 RNA 聚合酶 β 亚基侧瓣顶端的一个 α 螺旋区，在处于转录终止子位点的 RNA 中它可与暂停螺旋环相互作用。

fluor：荧石。被放射性发射源激活而发射光子的物质。

fluorescence resonance energy transfer（FRET）：荧光共振能量传递。一种用于测量两个分子或同一大分子两个组分之间距离的分析技术。其原理是两个彼此靠近的荧光分子之间有共振能量的转移，但转移效率随两个分子间距的增加而减弱。

fluorescent probe：荧光探针。在荧光共振能量传递实验中的一个荧光分子。

fMet：*N*-甲酰甲硫氨酸。见 *N*-formyl methionine。

Fos：AP-1 激活因子的两个亚基之一（另一个是 Jun）。

fragment reaction：片段化反应。用简单底物作为肽酰转移酶反应中的替代物。用与 fMet 连接的一个 $tRNA_f^{Met}$ 的 6 核苷酸片段替代肽酰-tRNA，用嘌呤霉素替代氨酰-tRNA。通过反应获得的产物是从核糖体释放的 fMet-嘌呤霉素。

frameshift mutation：移码突变。在基因编码区中由于一个或两个碱基的插入或缺失，改变了相应的 mRNA 可读框。

free radical：游离自由基。没有配对电子的非常活跃的化学物质，可以攻击和损伤 DNA。

functional genomics：功能基因组学。研究生物在不同时间和不同条件下全基因组的基因表达模式。

functional SELEX：基于功能的指数方式富集核酸配基的系统进化筛选技术（systematic evolution of ligands by exponential enrichment，SELEX）。一种基于功能（如被剪接的能力）富集核酸的 SEL-

EX 方法。

fusidic acid：梭链孢酸。一种在 GTP 水解后阻止 EF-G 从核糖体释放的抗生素，它将阻碍转位后的翻译过程。

fusion protein：融合蛋白。由含有两个融合在一起的可读框（ORF）（其中的一个或两个可能是不完整的）的重组 DNA 所表达的蛋白质。

G

G protein：G 蛋白。一类被 GTP 结合而激活，由自身 GTP 酶水解被结合的 GTP 为 GDP 而失活的蛋白质。

G-segment：G 片段。一个 DNA 片段，它能阻断拓扑异构酶 II 在表达活性时 T 片段穿行通道的形成。

galactoside permease：半乳糖苷透性酶。由大肠杆菌乳糖操纵子编码的、将乳糖输入到细胞内的酶。

galactoside transacetylase：半乳糖苷转乙酰酶。乳糖操纵子编码的 3 种酶之一，可使如乳糖中的半乳糖苷乙酰化，但它在乳糖操纵子中的重要性还不明了。

GAGA box：GAGA 框。果蝇绝缘子的元件。

GAL4：通过与上游调控元件（UAS_G）结合来激活酵母的半乳糖利用（GAL）基因的转录因子。

gamete：配子。单倍体性细胞。

gamma ray：γ 射线。能使细胞组分离子化的极高能量的射线。产生的离子可造成染色体的断裂。

GAP：GTP 酶激活因子蛋白。见 GTPase activator protein。

GC box：GC 框。一条链上的 GGGCGG 序列的六聚体，存在于许多哺乳动物的结构基因的启动子中，为转录因子 SP1 的结合位点。

GDPCP：在 β 和 γ 磷酸基团之间以亚甲基连接的一种不能水解的 GTP 类似物。

gel electrophoresis：凝胶电泳。一种在琼脂糖或聚丙烯酰胺胶上分离某种物质（一般为核苷酸或蛋白质）的电泳方法。

gel filtration：凝胶过滤。根据分子大小进行物质分离的柱状层析方法。由于小分子能进入凝胶珠孔而大分子不能进入，所以小分子比大分子需要更长的时间通过柱子。

gel mobility shift assay：凝胶阻滞分析、凝胶迁移变动分析。一种分析 DNA 与蛋白质之间相互结合的方法。将一段短的被标记的 DNA 与一种蛋白质混合后电泳，如果 DNA 结合到蛋白质上，它的电泳迁移率将会大大降低。

gene：基因。遗传的基本单元，含有生成一个 RNA 或在大多数情形下生成一个多肽的信息。

gene cloning：基因克隆。把一个基因插入到如细菌等生物中，通过与寄主一起复制产生该基因的多个拷贝。

gene cluster：基因簇。真核染色体上定位在一起的一组相关基因。

gene conversion：基因转换。基因的一种序列转换为该基因的另一种序列。

gene expression：基因表达。基因形成产物的过程。

general transcription factor：通用转录因子。与 RNA 聚合酶一起参与形成转录前起始复合体的真核生物蛋白。

genetic code：遗传密码。一套由密码子及其代表的氨基酸（或终止子）组成的 64 个密码。

genetic linkage：遗传连锁。同一染色体上基因间的物理关系。

genetic mapping：遗传作图。确定基因间的线性排列顺序和间距。

genetic marker：遗传标记。在基因组作图中，能够被用做标记点的一个突变基因或其他特殊性状。

genome：基因组。一个遗传体系中一套完整的遗传信息，如细菌的基因组就是其单个环形染色体。

genomic functional profiling：基因组功能图谱。确定一个有机体生命的所有阶段中所有基因的表达模式。

genomic library：基因组文库。一套直接来源于基因组而非 mRNA 的 DNA 片段克隆。

genomics：基因组学。研究全基因组的结构和功能的学科。

genotype：基因型。特定个体的等位基因组成。在一个二倍体个体中位点 *A* 的基因型可能是 *AA*、*Aa* 或 *aa*。

GG-NER：全基因组单核苷酸切补修复。见 globe genome NER。

gigabase pair（Gb）：10 亿碱基对。10 亿个碱基对。

G-less cassette：少 G 框。双链 DNA 片段的非模板链上缺乏 G。可将少 G 框置于启动子控制下来检测在体外 GTP 缺失时的转录。因为不需要 GTP，所以少 G 框能产生转录产物，但对于其他非特异性的转录，在缺少 GTP 时，只能产生很短的转录产物。

globe genome NER（GG-NER）：全基因组单核苷酸切补修复。可以在全基因组内移除损伤核苷酸的切补修复。

glucose：葡萄糖。常被许多生命形式作为能源的一种简单的六碳糖。

glutamine-rich domain：富含谷氨酰胺结构域。一个

富含谷氨酰胺的转录激活域。

glycosidic bond (in a nucleoside)：糖苷键（在核苷中）。在 RNA 或 DNA 中连接碱基和糖（核糖或脱氧核糖）的键。

Golgi apparatus：高尔基复合体。一种能包装新合成的蛋白质以便向细胞外输出的膜细胞器。

gp5：噬菌体 M13 基因 5 的产物。噬菌体单链 DNA 结合蛋白。

gp28：噬菌体 SP01 基因 28 的产物。噬菌体中期基因特异性 σ 因子。

gp32：噬菌体 T 基因 32 的产物。噬菌体单链 DNA 结合蛋白。

gp33 and gp34：噬菌体 SP01 基因 33 和 34 的产物。它们一起构成晚期基因特异性 σ 因子。

gpA：ϕX174 噬菌体基因 A 的产物。在噬菌体 DNA 复制中起重要作用，作为核酸酶使 RF 的一条链产生缺刻，再行使解旋酶功能打开母本双链 DNA。

GRAIL (gene recognition and assembly internet link)：基因识别及组装互联网。在数据库中鉴别基因的程序。

GRB2：一种接头蛋白，含有能识别一种信号转导蛋白上的磷酸酪氨酸的 SH2 结构域，以及另一个能识别其他信号转导蛋白上的富含脯氨酸螺旋的 SH3 结构域，从而传递信号。

g-RNA：向导 RNA。见 guide RNA (editing)。

group I intron：I 型内含子。由一个游离的鸟嘌呤核苷或鸟嘌呤核苷酸启动的自我剪接型内含子。

group II intron：II 型内含子。由形成索套的中间结构所启动的自我剪接型内含子。

GTPase activator protein (GAP)：GTP 酶激活因子蛋白。一种能够激活 G 蛋白的内源 GTP 酶活性从而导致 G 蛋白失活的蛋白质。

GTPase-associated site：GTP 酶相关位点。核糖体上与 G 蛋白起始、延伸和终止因子相互作用并激活其 GTP 酶活性的位点。

guanine (G)：鸟嘌呤。DNA 中与胞嘧啶配对的嘌呤碱基。

guanine nucleoside exchange protein：鸟嘌呤核苷交换蛋白。在 G 蛋白上用 GTP 替代 GDP 而激活 G 蛋白的一类蛋白质。

guanosine：鸟嘌呤核苷。一种含有鸟嘌呤碱基的核苷。

guide RNA (editing)：向导 RNA（编辑）。结合在 mRNA 前体不同区域上的小 RNA，作为编辑某个区域上游的模板。

guide sequence (splicing)：向导序列（剪接）。一段 RNA 区域，与 RNA 的其他区域结合以帮助剪接过程的正确定位。

H

hairpin：发夹。由一条单链 DNA 或 RNA 内的反向重复序列在分子内的碱基配对形成像发夹样的结构。

half-life：半衰期。分子群体中一半分子消失所耗的时间。

haploid：单倍体。配子中的染色体数目（n）。

haplotype：单倍体型。单条染色体上特定等位基因类型的组合。

haplotype map：单倍体型图。显示单倍体型位置的基因组图谱。

HAT：组氨酸乙酰转移酶。见 histone acetyl transferase。

HAT-A：组氨酸乙酰转移酶 A。一种能使核心组蛋白乙酰化并在基因调控中起作用的组氨酸乙酰转移酶。

HAT-B：组氨酸乙酰转移酶 B。一种能在组蛋白 H3 和 H4 组装配成核小体前发生乙酰化的组氨酸乙酰转移酶。

HCR：亚铁血红素控制阻遏蛋白。见 heme-controlled repressor。

HDAC1 and HDAC2：组蛋白去乙酰化酶 1 和组蛋白去乙酰化酶 2。两种组蛋白去乙酰化酶。

heat shock gene：热激基因。在包括热等环境协迫条件下开启以做出响应的基因。

heat shock response：热激响应。细胞对热或其他逆境损伤的响应。细胞启动编码分子伴侣的热激基因帮助变性蛋白重新折叠，并启动蛋白酶降解完全失活的蛋白质。

helicase：解旋酶。解开多聚核苷酸双螺旋的一种酶。

helix-loop-helix domain (HLH domain)：螺旋-环-螺旋结构域（HLH 域）。蛋白质的一种结构域，能够与另一个螺旋-环-螺旋域通过形成卷曲螺旋而构成二聚体。

helix-turn-helix：螺旋-转角-螺旋。某种 DNA 结合蛋白（特别是来自原核蛋白）的结构基序，能嵌于 DNA 的大沟内，并赋予蛋白质与 DNA 结合的能力和特异性。

helper virus (or phage)：协助病毒（或噬菌体）。为有缺陷的病毒提供相应功能补偿的另一种病毒，从而使缺陷病毒得以复制。

heme-controlled repressor (HCR)：亚铁血红素调控阻遏蛋白。使 eIF-2α 磷酸化的一种蛋白激酶，能增强它与 eIF-2B 的结合，从而阻断转录起始。

hemoglobin：血红蛋白。在红血细胞中红色携氧蛋白。

hereditary nonpolyposis colon cancer（HNPCC）：遗传性非息肉结肠癌。由于错配修复失败而导致的一种常见的人类遗传性结肠癌。

heterochromatin：异染色质。浓缩且无活性的染色质。

heteroduplex：异源双链核酸分子。两条链不完全互补的双链多聚核苷酸。

heterogeneous nuclear RNA（hnRNA）：核内不均一RNA。在细胞核内发现的一类大小不均一的RNA，包括未剪接的前体mRNA。

heteroschizomer：异裂酶。识别相同限制性位点，但切割位置不同的限制性内切核酸酶。

heterozygote：杂合子。在一个特定基因位点上具有两个不同形式等位基因的二倍体基因型，如A_1A_2。

histone acetyl transferase（HAT）：组蛋白乙酰转移酶。将乙酰CoA的乙酰基转移到组蛋白上的一种酶。

histone chaperone：组蛋白伴侣。将组蛋白负载到裸露DNA上形成核小体的蛋白质。

histone code：组蛋白密码。在对基因转录有特定效应的调控区邻近的核小体上存在的一套组蛋白修饰方式。

histone fold：组蛋白折叠。由两个环连接3个螺旋组成的一种组蛋白的结构基序。

histone methyl transferase（HMTase）：组蛋白甲基转移酶。含有能将甲基转移至核心组蛋白的染色质重建功能域的酶。

histone：组蛋白。存在于大多数真核染色体上能与DNA密切相连的5种小的碱性蛋白质。

HLH domain：螺旋-环-螺旋域。见helix-loop-helix domain。

HMG domain：HMG域。在一些结构转录因子上发现的一种类似于HMG结构域的区域。

HMG protein：高迁移率蛋白。具有高迁移率（高迁移率基团）的核蛋白。某些HMG蛋白已显示有调控转录的作用。

HNPCC：遗传性非息肉大肠癌。见hereditary non-polyposis colon cancer。

hnRNA：核内不均一RNA。见heterogeneous nuclear RNA。

hnRNP A1：异质核糖核蛋白A1。与ESS结合并协助抑制剪接的一种异质核糖核蛋白。

hnRNP protein：hnRNP蛋白。与hnRNA结合的蛋白质。

Holliday junction：Holliday联接。在重组过程中由第一链交换形成的分支DNA结构。

homeobox（HOX）：同源异型框。在同源异型基因和真核细胞的发育调控基因中发现的一段180bp的序列，编码一个同源异型结构域。

homeodomain（HD）：同源异型结构域。一个包含60个氨基酸，能与DNA结合的蛋白质结构域，可使蛋白质紧密结合于特异DNA区域。其结构和与DNA互作的模式类似于螺旋-转角-螺旋域。

homeotic gene：同源异型基因。发生突变后能导致身体的某部分转变为另一部分表型的基因。

homologous chromosome：同源染色体。除了等位基因差异和遗传组成外，其大小、形状都相同的染色体。

homologous（gene or protein）：同源（基因或蛋白）。源于进化相关性产生的相似性。

homologous recombination：同源重组。要求重组DNA分子间有广泛的序列相似性的重组。

homologs：同源基因。由一个共同祖先基因进化而来的基因，包括直系同源基因和旁系同源基因。

homozygote：纯合子。一个特定基因座上的两个等位基因都相同的二倍体基因型，如A_1A_1或aa。

hormone response element：激素应答元件。对结合于配体上的核受体产生响应的增强子。

housekeeping gene：看家基因。编码所有类型细胞基本生命过程所需蛋白质的基因。

HP1：一个与组蛋白甲基转移酶相关并且含有染色质重建功能域的蛋白质。

HTF island：HTF岛。见CpG island。

human endogenous retrovirus：人类内源反转录病毒。人类细胞中的一种转座缺陷且含有LTR的反转录转座子。

human immunodeficiency virus（HIV）：人类免疫缺陷病毒。导致获得性免疫缺陷综合征（AIDS）的反转录病毒。

HU protein：HU蛋白。一类诱导*ori*C弯曲从而促使开放复合体形成的、小的DNA结合蛋白。

hybrid dysgenesis：杂种不育。在果蝇中观察到的两个特定亲本的杂种后代因染色体损伤过多而不育的一种现象。

hybridization（of polynucleotide）：（多聚核苷酸的）杂交。两条来源不同的多聚核苷酸链形成的双链结构。

hybrid polynucleotide：杂合多聚核苷酸链。多聚核苷酸杂交的产物。

hydrogen bond network：氢键网络。两个及两个以上分子间的氢键网络。

hydroxyl radical：羟基自由基。含有未配对电子的羟基基团。它们是高度活跃的，并可攻击和打断DNA。因此成为足迹法的有效试剂。

hydroxyl radical cleavage analysis：羟基自由基裂解分析。见 hydroxyl radical probing。

hydroxyl radical probing：羟基自由基探测。一种探测蛋白质在 RNA 上结合位置的技术。将含铁离子（Fe^{3+}）的化合物连接到蛋白质的半胱氨酸上，然后使该蛋白质再结合于一个 RNA 或一个含 RNA 的复合体上。由该铁离子产生的羟基自由基可以打断邻近位点的 RNA，这种断裂可以通过引物延伸实验进行检测。

hyperchromic shift：增色转变。DNA 溶液变性时可表现在 260nm 处吸光值增加的反应。

I

identity element：认同元件。能被一个 DNA 结合域识别的碱基或其他 DNA 元件。

IF1：（翻译）起始因子 1。经过一轮翻译后能促使核糖体解离的原核生物翻译起始因子，也能增加其他两类翻译起始因子的活性。

IF2：（翻译）起始因子 2。负责将 fMet-$tMet_f^{Met}$ 结合到核糖体的原核生物翻译起始因子。

IF3：（翻译）起始因子 3。负责将 mRNA 结合到核糖体以及在完成一轮翻译后保持核糖体亚基处于分离状态的原核生物翻译起始因子。

immune（λ phage）：免疫（λ噬菌体）。一个 λ 噬菌体的溶源性细菌如果不能被第二种噬菌体感染，那么它对另一种 λ 噬菌体的超感染是免疫的。

immunity region：免疫区。λ 噬菌体或类似于 λ 噬菌体的调控区，包含阻遏物基因和被该阻遏物识别的操纵基因。

immunoblot：免疫印迹。见 Western blotting。

immunoglobulin（antibody）：免疫球蛋白（抗体）。一种能够非常特异地与入侵物质结合并警示个体的免疫防御系统摧毁入侵者的蛋白质。

immunoprecipitation：免疫沉淀反应。标记蛋白与一个特异抗体或抗血清作用，然后通过离心进行沉淀的技术。一般用电泳和放射性自显影检测被沉淀的蛋白质。

***in cis*：顺式排列。**两个基因同时位于同一条染色体上的状态。

in silico：计算机模拟。仅通过计算机（其中的硅芯）运行。

***in trans*：反式排列。**两个显性基因位于不同染色体上的状态。

incision：切割。用内切核酸酶切割 DNA 链。

inclusion body：包含体。大肠杆菌中由于外源基因高水平表达所形成的不溶性蛋白质集合体。它们通常是没有活性的，但常常可以通过变性和复性恢复活性。

independent assortment：独立分配。孟德尔发现的一个法则，认为在不同染色体上的基因是独立遗传的。

inducer：诱导物。一种解除操纵子负调控的物质。

initiation factor：起始因子。帮助催化翻译起始的一种蛋白质。

initiator：起始子。转录起始位点附近的一个位点，它对某些 II 类启动子的转录效率十分重要，尤其是缺少 TATA 盒的启动子。

in-line probing：串联探测。通过测量 RNA 被裂解的难易程度进而检测 RNA 中二级结构的方法。无结构的 RNA 更容易呈现反应物的串联排列，更易被裂解。

INO80：一种与 SW12/SNF2 同源的酵母核小体重建因子。

inosine（I）：次黄嘌呤核苷。含有与胞嘧啶碱基配对的次黄嘌呤的核苷。

insertion sequence（IS）：插入序列。细菌中发现的一种仅包含末端反向重复序列和转座所需基因的简单转座子。

***in situ* hybridization：原位杂交。**将被标记的探针直接与生物样本（如胚胎切面甚至是伸展状态的染色体）进行杂交，以寻找基因或基因的转录物。

insulator：绝缘子/隔离子。屏蔽一个基因的增强子正效应或沉默子负效应的 DNA 元件。

insulator body：绝缘/隔离小体。两个或多个绝缘子与其结合蛋白的复合体。

integrase：整合酶。能将一段核苷酸整合到另一段核苷酸中的酶。例如，整合一个反转录病毒的原病毒到宿主基因组中。

intensifying screen：增感屏。一种增强由放射性物质产生的放射自显影信号的屏。该屏包含能被放射性的发射源激活发射光子的氟类物质。

interactome：互作组。有机体中所有蛋白质间相互作用的总和。

intercalate：插入。插入到两个 DNA 碱基对之间。

interferon：干扰素。一种由双链 RNA 激活的在细胞中具有多种作用的抗病毒蛋白。

intergenic suppression：基因间抑制。一个突变基因抑制另一个突变基因的表型。

intermediate：中间物。生化途径中的一种底物-产物形式。

internal guide sequence：内部引导序列。具有自我剪

接内含子特征的核酶内的一个区域，能将 RNA 的其他部分置于被催化的位置。

internal ribosome entry sequence（IRES）：内部核糖体进入序列。能使核糖体无须从转录物 5′端扫描，而直接在转录物中部结合并翻译的序列。

intervening sequence（IVS）：间隔序列。见 intron。

intracistronic complementation：顺反子内互补作用。同一基因内的两个突变间的互补。通过不同缺陷单体之间的合作，可形成一个有活性的寡聚蛋白时发生。

intrinsic terminator：内在终止子。一种不需要其他终止因子（如 σ 因子）帮助的细菌终止子。

intron：内含子。打断基因转录部分的一段区域。内含子是被转录的，但在转录物成熟过程中将被剪接掉。这个词也指 DNA 及其 RNA 产物的间隔序列。

intron definition：内含子界定。剪接因子识别内含子末端的一种剪接模式。

intronic silencing element：内含子沉默元件。内含子中抑制剪接的区域。

inverted repeat：反向重复。DNA 的对称序列，DNA 双链中一条链的正向阅读和互补的反向阅读序列是相同的，如GGATCC

CCTAGG

IRE：金属反应元件。见 iron response element。

iron response element（IRE）：铁响应元件。mRNA 非翻译区的一个茎-环结构，与铁调节蛋白结合后反过来影响 mRNA 的寿命或翻译能力。

iron regulatory protein（IRP）：铁调节蛋白。与 IRE 结合的蛋白质。又见 aconitase。

IRP：见 iron regulatory protein。

isoaccepting species（of tRNA）：（tRNA 的）同工受体型。可以负载同样氨基酸的两个或多个tRNA。

isoelectric focusing：等电聚焦。根据 pH 梯度电泳蛋白质的混合物直到每种蛋白质都停滞在与它等电点相同的 pH。因为蛋白质在其等电点处无净电荷，因此它们不再向阳极或阴极移动。

isoelectric point：等电点。蛋白质无净电荷时的 pH。

isoschizomer：同裂酶。识别、切割相同限制性位点中相同位置的两种或多种限制性内切核酸酶。

ISWI：一类帮助染色质重建的辅激活因子家族。

J

joining region（J）：连接区。编码可变区最后 13 个氨基酸的一段免疫球蛋白基因序列。多个连接区中的一个区域通过染色体重排被连接到可变区的其余部分，从而产生更多的基因多样性。

joint molecule：接合分子。大肠杆菌同源重组过程中联会后期的中间产物，此时已开始链的交换，两条 DNA 链互相缠绕在一起。

Jun：AP-1 激活因子的两个亚基之一（另一个是 Fos）。

K

keto tautomer：酮基互变异构体。在核苷酸中发现的尿嘧啶、胸腺嘧啶和鸟嘌呤的正常互变体。

kilobase pair（kb）：千碱基对。1000 个碱基对。

kinetic experiment：动力学实验。测量反应动力学（速度）的实验。由于化学反应在很短时间内发生，因而要求快速测量。

kinetoplast：动基体。锥体虫的线粒体。由许多微环和大环基因组成。

Klenow fragment：Klenow 片段。DNA 聚合酶 I 通过蛋白酶水解得到的一个片段，缺少母体酶的5′→3′外切核酸酶活性。

known gene：已知基因。从基因组测序计划得到的与前期已鉴定过的基因一致的基因。

Kozak's rule：Kozak 法则。真核翻译起始信号附近保证准确翻译需要的一组最佳核苷酸序列。Kozak 法则最重要的是与 AUG 起始密码子相关的－3位碱基应该是嘌呤，最好是腺嘌呤，＋4 位碱基应该是 G。

Ku：DNA-PK 中含 ATP 酶的调节亚基。与由染色体断裂形成的双链 DNA 末端结合，在末端连接发生之前起保护作用。

L

L1：一种在人类大量存在的 LINE 因子，至少有 100 000 拷贝，约占人类基因组的 15%。

***lacA*：**编码半乳糖乙酰转移酶的大肠杆菌基因。

***lacI*：**编码乳糖操纵子阻遏物的大肠杆菌基因。

***lac* operon：乳糖操纵子。**编码细胞乳糖代谢酶的操纵子。

***lac* repressor：乳糖阻遏物。**大肠杆菌 *lacI* 基因产物，该蛋白质形成一个四聚体结合到乳糖操纵子上，从而阻遏乳糖操纵子。

lactose：乳糖。由半乳糖和葡萄糖两个单糖组成的双糖。

***lacY*：**编码半乳糖透性酶的大肠杆菌基因。

***lacZ*：**编码 β-半乳糖苷酶的大肠杆菌基因。

lagging strand：后随链。DNA 半保留复制模型中分段式合成的链。

large T antigen：大 T 抗原。SV40 病毒早期区域的主要产物。一种与病毒 *ori* 位点结合并解旋 DNA

链以便于引物合成的 DNA 解旋酶。也可以导致哺乳动物细胞的恶性转化。

lariat：套索。命名某些剪接反应中形成的套索形中间产物。

leader：前导区。mRNA 5′端的一段不翻译的碱基序列（5′非翻译区）。

leading strand：前导链。DNA 半保留复制模型中连续性合成的链。

LEF-1：淋巴样增强子结合因子，一种结构性转录因子。

leucine zipper：亮氨酸拉链。一种 DNA 结合蛋白的结构域，其中几个亮氨酸按一定的间隔规律排列。这种结构有利于与另一个亮氨酸拉链形成二聚体。此二聚体可与 DNA 特异性结合。

LexA：大肠杆菌 *lexA* 的基因产物。作用之一是 *umuDC* 操纵子的阻遏物。

light repair：光修复。见 photoreactivation。

LIM homeodomain（LIM-HD）activator：LIM 同源域激活因子。与 CLIM 辅激活因子和 RLIM 辅阻遏物相关的激活因子。

limited proteolysis：有限蛋白酶解。用蛋白水解酶温和处理，使蛋白质分解为各结构域的组分。

LINE：长散布元件。见 long interspersed element。

linker scanning mutagenesis：接头分区诱变。利用合成的双链寡核苷酸（接头）替换天然 DNA 中的小片段（大约 10bp）而产生的成簇突变。

liposome：脂质体。一种被包裹的小泡。用以介导 DNA 进入细胞。

liquid scintillation counting：液体闪烁计数。通过一种闪烁的液体包围某种物质来测定此物质的放射性的技术。这种闪烁液含有氟类物能够被具有放射性的发射源激活而发射光子。

locus（loci，pl.）：基因座（复数 loci）。基因在染色体上的位置，通常与基因同义。

locus control region（LCR）：基因座控制区。一段染色质区域，如球蛋白基因相关区域，它可确保相关基因的活性，而这与染色体的位置无关。

long interspersed element（LINE）：长散布元件。在哺乳动物中最丰富的非 LTR 反转录转座子。

long terminal repeat（LTR）：长末端重复序列。在反转录病毒的原病毒或含 LTR 反转录转座子的两端发现的长几百个 DNA 碱基对的区域。

looping out：环凸。结合在一个 DNA 分子的相距较远位点上的 DNA，结合蛋白同时发生相互作用的过程，可引起两个位点间的 DNA 形成一个环。

LTR：长末端重复序列。见 long terminal repeat。

LTR-containing retrotransposon：含 LTR 的反转录转座子。两端含有 LTR 的反转录转座子。除了不涉及传染性病毒外，其复制方式与反转录病毒相同。

luciferase：萤光素酶。将萤光素转化为可发光并容易检测的化学发光的酶。萤火虫的萤光素酶基因常在真核细胞转录和翻译实验中作为报告基因。

luxury gene：奢侈基因。编码特殊细胞产物的基因。

lysis：裂解。使细胞膜破裂，如烈性噬菌体所引起的。

lysogen：溶源体。带有原噬菌体的细菌。

M

Mad-Max：哺乳动物的一个阻遏物。

MAPK：分裂素激活的蛋白激酶。见 mitogen activated protein kinase。

mariner：人类的一种缺陷型转座子，仅依据以前的推测，它可直接通过 DNA 复制进行转座。

marker：标记。在基因组的已知位点上作为标识的基因或突变。

Maskin：一种非洲爪蟾蛋白，它通过与 CPEB 和 eIF4E 的结合抑制细胞周期蛋白 B 的 mRNA 翻译，并阻止 eIF4E 与 eIF4G 结合。当 CPEB 被磷酸化而离开复合体后，Maskin 就释放出 eIF4E，随后启动翻译。

mass spectrometry：质谱分析。使分子发生离子化并将其发射至标靶上的一种高分辨率分析技术。假定其均为带有电荷的单体分子，则它们飞行至标靶所需时间与其质量相关。质量提供了有关分子性质的重要信息。

maternal gene：母本基因。母体卵子发生过程中所表达的基因。

maternal message：母本信息。受精前卵母细胞产生的 mRNA。许多母本信息在受精前保持着不翻译状态。

maternal mRNA：见 maternal message。

maxicircle：大环。动基体中发现的 20～40kb 的环状 DNA。含有基因（隐秘基因）和编码某些动基体的 gRNA。

mediator：中介因子。一种与激活因子结合并能帮助激活因子促进前起始复合体组装的酵母辅激活因子。

megabase pair（Mb）：兆碱基对。100 万碱基对。

meiosis：减数分裂。产生只含亲本细胞一半染色体数目的配子（或孢子）的细胞分裂。

MEK（MAPK/ERK kinase）：MAPK/ERK 激酶。一种由 Raf 激活信号转导丝氨酸/苏氨酸蛋白激酶，并通过磷酸化激活 ERK 的酶。

Mendelian genetics：孟德尔遗传学。见 transmission

genetics。

merodiploid：部分二倍体。只在某些基因上呈二倍化的细菌。

message：信息。见 mRNA。

messager RNA：信使 RNA。见 mRNA。

methylation interference assay：甲基化干扰分析。一种检测能与特殊蛋白质发生相互作用的 DNA 位置的重要方法，这种 DNA 甲基化位点干扰其与蛋白质的结合。

micrococcal nuclease：微球菌核酸酶。一种能降解核小体之间 DNA 的核酸酶，仅留下核小体 DNA 的酶。

microRNA（miRNA）：微小 RNA（miRNA）。一种在细胞中自然产生的 18～25nt 的小 RNA，可通过使特定的 mRNA 降解或阻止其翻译来控制靶基因的表达。

microsatelite：微卫星。一种短的串联重复多次的 DNA 序列（通常 2～4bp）。特定的微卫星以不同的长度散布在真核生物基因组中。

minicircle：微环。在动基体中发现的 1～3kb 环状 DNA，编码动基体中的一些 gRNA。

minimal genome：最小基因组。维持有机体生命的最少的基因集合。

minisatellite：小卫星。一种通常 12bp 或更长，且串联重复多次的短序列。

minus ten box（－10 box）：－10 框。位于大肠杆菌基因启动子元件中心转录起始位点上游约 10bp 的一段保守序列。

minus thirty-five box（－35 box）：－35 框。位于大肠杆菌基因启动子元件中心转录起始位点上游约 35bp 的一段保守序列。

miRISC：果蝇中与 miRNA 一起参与基因表达调控的 RISC。

mismatch repair：错配修复。在随机情况下尤其是对在编辑过程中渗入到新合成 DNA 中的错配碱基的修复机制。

missense mutation：错义突变。密码子中一个碱基的改变导致相应蛋白质氨基酸的改变。

mitogen：分裂素。一种能促进细胞分裂的物质，如激素或生长因子。

mitogen activated protein kinase（MAPK）：受分裂素激活的蛋白激酶。由生长因子等分裂素启动的信号转导导致被磷酸化激活的一种蛋白激酶。

mitosis：有丝分裂。产生与含有亲代细胞相同核物质的两个子代细胞的细胞分裂。

model organism：模式生物。替代人类用以研究的生物。选取条件如下：基因组较小、生长周期短，容易进行基因操作。

molecular chaperone：分子伴侣。见 chaperones。

M1 RNA：RNase P 的 RNA 催化亚基。

mRNA（messenger RNA）：信使 RNA。携带一种或多种蛋白质遗传信息的转录物。

Mud2p：酵母的剪接因子，它能与 3′剪接点和分支点桥结蛋白结合（BBP），从而可以锚定外显子-内含子界限内的 3′剪接位点。

multiple cloning site（MCS）：多克隆位点。在克隆载体中的一段随机分布着多个限制性内切核酸酶位点的区域，此区内可插入外源 DNA。

mutagen：诱变剂。可诱导有机体发生突变的物质。

mutant：突变体。至少在一处发生突变的生物体或遗传体系。

mutation：突变。由 DNA 的碱基改变或染色体变异造成的遗传变异的最初来源。自发突变往往无法解释，而诱发突变则是由于特殊的诱变剂引起。

mutator mutant：增变突变体。突变累积速度快于野生型的突变体。

N

N：编码抗终止蛋白 N 的 λ 噬菌体基因。

N：N 基因产物，抑制定位在 λ 噬菌体早期基因之后的转录终止的抗终止蛋白。

N utilization site（*nut* site）：N 利用位点（*nut* 位点）。λ 噬菌体早期基因上的一个位点，可使 N 发挥其抗终止的作用。*nut* 位点的转录可提高相应的 N 的转录。N 可与能结合 RNA 聚合酶的多种蛋白质相互作用，从而使聚合酶在早期基因后忽视终止信号而表现转录通读。

NAS：见 nonsense-associated altered splicing。

NC2：见 negative cofactor 2。

N-CoR/SMRT：哺乳动物中与核受体协同作用的辅阻遏物。

ncRNA：见 noncoding RNA。

negative cofactor 2（NC2）：负辅因子 2。刺激含 DPE 的启动子转录，但抑制含 TATA 盒的启动子转录的蛋白质。

negative control：负调控。一种调控系统，当调控蛋白（如阻遏物）被去除后基因的表达才能开启。

neoschizomer：见 heteroschizomer。

NER：见 nucleotide excision repair。

***Neurospora crassa*：粗糙脉孢菌。**一种常见的面包霉菌。由 Beadle 和 Tatum 培育出来，可用于遗传研究。

NF-κB：见 nuclear factor kappa B。

***N*-formyl methionine（fMet）：*N*-甲酰甲硫氨酸。**细

菌翻译的起始氨基酸。

nick：缺口。单链 DNA 上的一个断裂。

nick translation：缺口平移。DNA 聚合酶在对 DNA 缺口下游的链进行降解的同时，在缺口后合成新链的过程。结果导致了缺口向 DNA 链 3′端方向发生了移动。

nitrocellulose：硝酸纤维素。一种经化学处理发生改变从而可吸附单链 DNA 和蛋白质的纸。用于与标记探针杂交前的 DNA 吸附和与抗体反应前蛋白质的吸附。

NMD：见 nonsense-mediated mRNA decay。

node：结点。共环体中两环的相互交叉点。

nonautonomous retrotransposon：非自主型的反转录转座子。一种不具长末端重复序列的转座子，不能编码蛋白质，所以需依赖其他反转录转座子进行转座。

noncoding base：非编码碱基。一种不能与自然界中的任何碱基正确配对的 DNA 碱基（如 3-甲基鸟嘌呤）。

noncoding RNA（ncRNA）：非编码 RNA（ncRNA）。不编码蛋白质的转录物。又见 transcripts of unknown function（未知功能转录产物）。

nonhomologous end-joining：非同源部的连接。真核生物中一种修复双链 DNA 断裂（染色体断裂）的机制。

non-LTR retrotransposon：无-LTR 反转录转座子。一种缺少长末端重复序列（LTR）的反转录转座子，复制机制不同于有 LTR 的反转录转座子。

nonpermissive condition：非许可条件。条件突变基因产物在这些条件下不能发挥作用。

nonreplicative transposition：非复制型转座。转座子在从 DNA 的一个位置转座到另一位置的过程中不进行复制的一种转座模式，转座后在原来的位置无相应的拷贝存在。

nonsense codons UAG，UAA，and UGA：无义密码子 UAG、UAA 和 UGA。这些密码子使核糖体停止合成蛋白质。

nonsense-associated altered splicing（NAS）：无义相关选择性剪接。真核生物对提前出现终止密码的处理系统。当提前出现的终止密码与起始密码位于同一可读框的现象被检测出时，NAS 系统就会激活选择性剪接程序，除去包含提前出现终止密码的前体 mRNA 部分。

nonsense-mediated mRNA decay（NMD）：无义密码子介导的 mRNA 降解。真核生物降解提前出现终止密码的系统。当在 mRNA 中出现一个下游去稳定元件与一个无义密码子之间的距离过长，细胞就会把无义密码子识别为提前出现的终止密码而将其 mRNA 降解。

nonsense mutation：无义突变。在基因编码区出现的提前终止密码突变。包括琥珀突变（UAG）、赭石突变（UAA）和乳白突变（UGA）。

non-stop mRNA：无终止 mRNA。没有终止密码子的 mRNA。

nontemplate DNA strand：非模板 DNA 链。与模板链互补的链。有时称编码链或有义链。

nontranscribed spacer（NTS）：非转录间隔。位于两个前体 rRNA 基因之间的 DNA 区域。

Northern blot：Northern 印迹。将 RNA 片段转移到支持介质的过程（见 Southern 印迹）。

nr：BLAST 搜索获得的默认（非丰余）的数据库。

nt：见 nucleotide。

nTAF（neural TAF）：神经系统的 TAF。与 TRF1 相关的 TAF。

α-NTD：细菌 RNA 聚合酶 α 亚基的氨基端结构域。

nuclear factor kappa B（NF-κB）：核因子 κB（NF-κB）。哺乳动物的激活因子，与其他因子，如 HMG（Y）协同作用，激活免疫系统中的 β 干扰素和其他基因。

nuclear receptor：核受体。与激素（如性激素、糖皮质激素、甲状腺激素）或其他的物质（如维生素 D 或视黄酸）相互作用并结合增强子促进转录的蛋白质。一些核受体存在于核内，一些在胞质中与配体结合形成复合物，然后转移到核内进而促进转录。

nucleic acid：核酸。由核苷酸相连组成的链状分子（DNA 或 RNA）。

nucleocapsid：核壳。一种由病毒基因组（DNA 或 RNA）及外壳蛋白组成的结构。

nucleolus：核仁。细胞核中发现的细胞器，在细胞分裂的某些过程中消失，含有 rRNA 基因。

nucleoside：核苷。与糖结合的碱基，包括核糖和脱氧核糖。

nucleosome：核小体。真核生物染色体中重复结构的基本单元，由 8 个组蛋白形成中心，外围缠绕 200bp 的 DNA 片段以及结合在组蛋白八聚体外的 1 分子组蛋白 H1 组成。

nucleosome core particle：核小体核心颗粒。核酶消化后剩余的 145bp 的核小体 DNA，包括中心的组蛋白八聚体，但不含组蛋白 H1。

nucleosome positioning：核小体定位。确定与基因启动子位置相关的核小体的特定位置。

nucleotide（nt）：核苷酸。DNA 或 RNA 链中的基本单元，由糖、碱基及至少一个磷酸基团组成。

nucleotide excision repair：核苷切补修复。在切补修复途径中，酶从损伤碱基的任一端切开 DNA 链，切除含有损伤的寡核苷酸。形成的缺口由 DNA 聚合酶和 DNA 连接酶填补。

NuRD：具有核心组蛋白去乙酰酶活性的核小体重建因子。

NusA：通过促进暂停在终止子处的转录物上形成发夹结构进而引发转录终止的一种细菌蛋白质。

O

O6-methylguanine methyltransferase：O6-甲基鸟嘌呤甲基转移酶。从烷基化 DNA 上接受甲基或乙基从而逆转 DNA 损伤的自杀性酶。

ochre codon：赭石密码子。一种终止密码 UAA。

ochre mutation：赭石突变。见 nonsense mutation。

ochre suppressor：赭石突变抑制子。一种 tRNA 带有能识别赭石密码子（UAA）的反密码子从而可抑制赭石突变。

Okazaki fragment：冈崎片段。后随链不连续合成形成的 1000～2000bp 的小 DNA 片段。

oligo（dT）cellulose affinity chromatography：oligo（dT）纤维素亲和层析。一种通过相对高离子强度缓冲液的作用使 poly（A）$^+$ RNA 结合到 oligo（dT）纤维素上，然后用水洗脱来纯化 poly（A）$^+$ RNA 的方法。

oligomeric protein：低聚蛋白。至少包含一个多肽亚基的蛋白质。

oligonucleotide：寡核苷酸。短 RNA 或 DNA 片段。

oligonucleotide array：寡核苷酸阵列。见 DNA microchip。

oncogene：原癌基因。一种表达后会导致细胞产生癌变的基因。

one gene-one polypeptide hypothesis：一个基因一个多肽假说。已被普遍接受的一种假说，即一个基因编码一种多肽。

oocyte 5S rRNA gene：卵母细胞 5S rRNA 基因。仅在卵母细胞中表达的 5S rRNA 基因（在非洲爪蟾单倍体中的数量约为 19 500 拷贝）。

opal codon：卵白密码子。一种终止密码子，UGA。

opal mutation：卵白突变。见 nonsense mutation。

opal suppressor：卵白突变抑制子。一种 tRNA 含有能识别卵白密码子（UGA）的反密码子从而可抑制卵白突变。

open complex：开放复合体。由 dnaA 蛋白和 *oriC* 组成的复合体，在 *oriC* 中的 3 个由 13 个碱基形成的区段已处于解链状态。

open promoter complex：开放型启动子复合体。由 RNA 聚合酶和原核生物启动子间紧密结合形成的复合体。“开放”的意思是约有 10bp 的 DNA 双链打开或分开。

open reading frame（ORF）：可读框。不含转录终止密码子的读码框。

operator：操纵基因。原核生物中与特异阻遏物紧密结合从而调节相邻基因表达的 DNA 元件。

operator constitutive mutation：操纵基因组成型突变。不能有效地与阻遏物结合的操纵基因的突变，导致该操纵子的表现型是组成型的，始终保持活性状态。

operon：操纵子。受操纵基因协同调控的一组基因。

O region：O 区域。*attP* 和 *attB* 间的一小段同源区域。

ORF：见 open reading frame。

***oriC*：**大肠杆菌复制起始位点。

origin of replication：复制起始位点。在复制子中复制起始的唯一位点。

ortholog：直系同源。由共同祖先的基因进化而来的不同生物的同源基因。

oxoG：见 8-oxoguanine。

oxoG repair enzyme：8-羟基鸟嘌呤修复酶。切割连接 8-羟基鸟嘌呤与脱氧核糖上的糖苷键的酶。

P

P_{RE}：在建立溶源状态过程中，转录阻遏蛋白基因所需的 λ 启动子。

P_{RM}：在维持溶源状态中，阻遏蛋白基因转录所需的 λ 启动子。

P site（ribosomal）：核糖体 P 位点。当新的氨酰-tRNA 进入核糖体时，一个缩合的肽酰-tRNA 结合的位点。

PAB I：见 poly(A)-binding protein I。

PAB II：见 poly(A)-binding protein II。

palindrome：回文序列。见 inverted repeat。

panediting：泛编辑。前体 mRNA 的泛编辑。

paper chromatography：纸色层析法。一种基于不同分子与纸及流动相溶剂的亲和力进行分离的色谱分析方法。

paralog：旁系同源基因。种内由基因复制进化而产生的同源基因。

paromomycin：巴龙霉素。一种能结合到核糖体 A 位点并降低翻译准确性的抗生素。

paranemic double helix：平行双螺旋。双链不相互缠绕的双螺旋，仅仅是一条链位于另一条链上，无须解旋就可分开。

passenger strand：随从链。当双链 siRNA 中的引导

链与 RISC 复合体结合后被弃置的那条链。

pathway (biochemical)：途径（生化）。一系列生物化学反应，其中一个反应的产物（中间产物）成为下一个反应的底物。

pause site：暂停位点。RNA 聚合酶在发生转录通读之前所暂停的 DNA 位点。

PAZ：Argonaute 蛋白上与单链 siRNA 结合的功能域。

pBR322：基因克隆的质粒载体之一。

PBS：见 primer-binding site。

PCNA：增殖细胞核抗原。在前导链合成过程中，保证 DNA 聚合酶 δ 的进行性特征的真核蛋白质。

PCR：见 polymerase chain reaction。

P element：P 元件。果蝇中控制杂交不育的转座元件，可用于诱导果蝇突变。

peptide bond：肽键。蛋白质中连接氨基酸的键。

peptidyl transferase：肽酰基转移酶。该酶是核糖体大亚基的组成部分，可催化蛋白质合成中的肽键形成。

permissive condition：许可条件。允许条件突变基因产物发挥功能的条件。

phage：噬菌体。细菌病毒。

phage 434：噬菌体 434。具有自身独特免疫区的 λ 样噬菌体。

phage display：噬菌体展示。当外源基因和噬菌体的衣壳蛋白基因以融合蛋白形式表达后，这一外源蛋白会展现在噬菌体的表面。

phagemid：噬菌粒。一种带有单链噬菌体复制起点的质粒克隆载体，在噬菌体感染时能产生单链 DNA 克隆。

phage P1：噬菌体 P1。用于克隆大片段 DNA 的裂解性大肠杆菌噬菌体。

phage P22：噬菌体 P22。具有独特免疫区的 λ 样噬菌体。

phage T7：噬菌体 T7。一种相对简单的大肠杆菌 DNA 噬菌体，与 T3 噬菌体属同一组。能编码相应大肠杆菌噬菌体的单亚基 RNA 聚合酶。

pharmacogenomics：药物基因组学。用患者的 SNP 预测其对各种药物的反应，使治疗更加个性化。

phenotype：表型。有机体在形态、生物化学、行为或其他方面表现的特征。通常只考虑某个特殊性状，如重量。

phorbol ester：弗波酯。一种可以通过级联反应（包括激活 AP1）以促进细胞分裂的化学物质。

phosphodiester bond：磷酸二酯键。核酸中连接核苷酸的糖-磷酸键。

phosphorimager：磷屏成像仪。实现磷屏成像的装置。

phosphorimaging：磷屏成像。一种无须使用胶卷而能自动测量底物（如印迹膜上）放射性强度的技术。

photoreactivating enzyme：光复活酶。见 DNA photolyase。

photoreactivation：光复活。通过 DNA 光解酶对嘧啶二聚体进行直接修复的过程。

physical map：物理图谱。基于 DNA 的物理特征，如限制性位点，而非基因位置所制作的遗传图谱。

PIWI：Argonaute 蛋白上具有切割靶 mRNA 活性的一个功能域。

plaque：噬菌斑。病毒在侵染宿主平皿上由于杀死细胞或减慢其生长而形成的空斑。

plaque assay：噬菌斑测试。通过测定经过一定稀释后的病毒所产生的噬菌斑数量来确定病毒（或噬菌体）浓度的一种测试方法。

plaque-forming unit (pfu)：噬菌斑形成单位 (pfu)。在噬菌斑测试中，病毒形成一个噬菌斑的能力。

plaque hybridization：噬菌斑杂交。筛选含目标基因的噬菌体克隆的，即用标记的探针与取自大量噬菌斑的 DNA 中的目标基因杂交进行检测。

plasmid：质粒。一种不依赖细胞染色体而独立复制的环状 DNA 分子。

plectonemic double helix：缠绕双螺旋。两条链相互缠绕并且只有通过解旋才能分开的一种双螺旋，如沃森-克里克式双螺旋。

point mutation：点突变。一个或几个相连碱基的改变。

poly(A)：多聚腺苷酸。在一个典型真核mRNA末端的大约 200 个串联的腺苷酸。

poly(A)-binding protein I (PAB I, Pab1p)：poly(A) 结合蛋白 I。结合在 mRNA 末端并有助于 mRNA 翻译的蛋白质。

poly(A)-binding protein II (PAB II)：poly(A) 结合蛋白 II。结合在前体 mRNA 末端新合成的 poly (A) 上并促使其延长的蛋白质。

polyadenylation：多聚腺苷酸化。将 poly (A) 加到 RNA 的 3′端。

polyadenylation signal：多聚腺苷酸化信号。控制转录物切割和腺苷酸化的一组 RNA 序列。典型的切割信号是：一段 AAUAAA 序列紧随 20～30 个富 GU 区，后跟富 U 区。切割后的 AAUAAA 序列就成为多聚腺苷酸化信号。

poly(A) polymerase(PAP)：poly(A) 聚合酶。将 poly(A)加到 mRNA 末端或其前体上的酶。

poly(A)$^+$ RNA：在 3′端含 poly(A) 的 RNA。

poly(A)⁻ RNA：在3′端不含poly(A)的RNA。

polycistronic message：**多顺反子信使**。一个携带多个基因信息的mRNA。

polymerase chain reaction（PCR）：**聚合酶链反应**。利用DNA区域的两端引物和DNA聚合酶的重复反应，实现对DNA片段进行扩增的过程。

polynucleotide：**多核苷酸**。以核苷为组成单元的多聚物，如DNA或RNA。

polypeptide：**多肽**。单一蛋白质链。

polyprotein：**多聚蛋白质**。能被加工成两个或多个小的、有功能多肽的一条长多肽，如反转录病毒中的pol多聚蛋白。

polyribosome：**多核糖体**。见polysome。

polysome：**多核糖体**。一条信使RNA结合（或被）多个核糖体（翻译）的现象。

positional cloning：**图位克隆**。对某一特定遗传性状的基因进行定位。

positive control：**正调控**。依赖于正向效应物如CAP（和cAMP）的基因表达的调控系统。

positive strand：**正链**。病毒基因组DNA中与其mRNA有相同遗传信息的链。

positive strand phage（or virus）：**正链噬菌体（或病毒）**。全基因组只有一条mRNA的RNA噬菌体（或病毒）。

postsynapsis：**联会后期**。大肠杆菌中同源重组的一个时期，单链DNA取代双链DNA中的一条链形成一条新的双螺旋。

postreplication repair：**复制后修复**。见recombination repair。

posttranscriptional control：**转录后调控**。在转录后阶段对转录物进行剪接、切割及修饰加工的基因表达调控。

posttranscriptional gene silencing（PTGS）：**转录后基因沉默**。见RNA interference。

posttranslational modification：**翻译后修饰**。蛋白质合成后发生的一系列改变。

Pot1p：裂殖酵母中结合于端粒末端并保护其免受降解和DNA修复酶作用的蛋白质。

ppGpp：预警信号因子3′-二磷酸和5′-二磷酸鸟嘌呤核苷（4磷酸鸟苷酸），当细菌受到饥饿胁迫时可使翻译速度降低。

predicted gene：**预测基因**。通过对基因组序列的比较，鉴定与EST序列有同源性的基因。

preinitiation complex：**前起始复合体**。转录开始前由RNA聚合酶和通用转录因子在启动子上组装形成的复合体。

presynapsis：**前联会期**。大肠杆菌中同源重组的一个时期，即在DNA双链侵入前由RecA（和SSB）覆盖在单链DNA上。

Pribnow box：**Pribnow盒**。见minus ten box（−10 box）。

primary structure：**一级结构**。多肽中氨基酸或DNA/RNA上的核苷酸的序列。

primary transcript：**初级转录物**。一个基因初始的、未加工的RNA产物。

primase：**引发酶**。引发体中用以合成引物的酶。

primer：**引物**。一小段RNA，可为DNA复制起始提供所需要的自由末端。

primer-binding site（PBS）：**引物结合位点**。反转录病毒RNA上的位点，使tRNA引物结合并起始反转录。

primer extension：**引物延伸**。对样品中转录物进行定量，同时也可定位转录物5′端的方法。具体方法为：将一段被标记的DNA引物与混合物中特定的mRNA杂交，用反转录酶将转录物延伸至5′端，然后将DNA产物进行电泳确定其大小和丰度。

primosome：**引发体**。大肠杆菌中由约20个多肽形成的复合物，可产生DNA复制所需的引物。

probe（nucleic acid）：**探针（核酸）**。一段被示踪物（通常为放射性的）标记的核酸，使研究者能够追踪探针与未知DNA之间的杂交。例如，放射性探针可用于鉴定电泳后的未知DNA条带。

processed pseudogene：**被加工型假基因**。一种由类反转座子活性所产生的假基因：经历了正常基因的转录、转录物的加工、反转录以及重新插入基因组的过程。

processing（of RNA）：**RNA加工**。发生在前体RNA成熟过程中一组剪切事件包括：剪接、5′端或3′端的修剪或从大的前体上剪切出rRNA。

processivity：**进行性**。一种酶能够持续结合在一种或多种底物上进行重复催化过程的趋势。因此，DNA聚合酶或RNA聚合酶不从模板上解离并持续进行产物合成的时间越长，其进行性就越强。

prokaryote：**原核生物**。无细胞核的微生物。由细菌包括蓝细菌（蓝-绿藻）和古菌组成。

proliferating cell nuclear antigen（PCNA）：**增殖细胞核抗原（PCNA）**。一种与真核生物DNA聚合酶δ结合的蛋白质，可增加其进行性。

proline-rich domain：**富含脯氨酸结构域**。富含脯氨酸的转录激活结构域。

promoter：**启动子**。位于基因转录起始位点上游的一段DNA序列，通常在转录起始前RNA聚合酶能与之结合。

promoter clearance：启动子清除。转录起始后 RNA 聚合酶离开启动子的过程。

proofreading（aminoacyl-tRNA synthetase）：校正（氨酰 tRNA 合成酶）。氨酰腺苷酸（有时是氨酰 tRNA）被水解的过程，如果它们所携带的氨基酸对合成酶来说太小的话。

proofreading（DNA）：校正（DNA）。细胞对 DNA 复制准确性进行检验并用正确碱基替换错配碱基的过程。

proofreading（protein synthesis）：校正（蛋白质合成）。在氨基酸被整合进合成中的蛋白质链之前所进行的对核糖体上氨酰 tRNA 正确性双重检查。

prophage：原噬菌体。整合到宿主染色体上的噬菌体基因组。

protease：蛋白酶。剪切蛋白质的酶。例如，将反转录病毒的多聚蛋白切割为功能性部分的酶。

proteasome：蛋白酶体。水解泛素化蛋白的一组蛋白质复合体（具有 26S 的沉淀系数）。

protein：蛋白质。由氨基酸亚单位组成的一种多聚体或多肽，有时蛋白质一词指超过一条且有功能的多肽链（如血红蛋白由 4 条多肽链组成）的多聚体。

protein footprinting：蛋白质足迹。一种类似于 DNA 酶足迹法的检测蛋白质之间结合位点的方法。再用蛋白酶温和消化。如果结合蛋白保护了标记蛋白的部分区域免受酶解，那么电泳酶解产物时就会缺少代表相应区域的条带。

protein kinase A（PKA）：蛋白激酶 A（PKA）。一种被 cAMP 激活的丝氨酸-苏氨酸特异蛋白激酶。

protein sequencing：蛋白质测序。测定蛋白质中氨基酸的组成顺序的过程。

proteolytic processing：蛋白质水解加工。将蛋白质剪切成片段。

proteome：蛋白质组。有机体一生中所能合成的所有蛋白质的结构和功能。

proteomics：蛋白质组学。研究蛋白质组的科学。

provirus：原病毒。插入宿主基因组的反转录病毒 RNA 产生的双链 DNA 拷贝。

Prp28：U5 snRNP 的蛋白质组分，对于 U6 在 5′剪接位点处替代 U1 的过程是必需的。

pseudogene：假基因。正常基因的非等位拷贝，由于发生变异而不能发挥功能。

pseudouridine：假尿苷。在 tRNA 中发现的一种核苷，其核糖被连接在 5 位 C 原子上，而非正常尿嘧啶碱基的 1 位 N 原子上。

p300：CBP 的同源物。

pUC vector：pUC 载体。基于 pBR322 质粒改造的载体，包括一个氨苄青霉素抗性基因和一个位于 *lacZ′* 基因中间的多克隆位点，可以对插入片段进行蓝/白斑筛选。

pulse-chase：脉冲追踪（技术）。先将一个物质（如 RNA）与一放射性前体进行短暂的或“脉冲”作用，以使其带上放射性，然后加入过量的未标记的前体来追踪此物质中的放射性。

pulsed-field gel electrophoresis（PFGE）：脉冲场电泳（PFGE）。一种反复交互变换电场方向的电泳技术，可分离大至几个 Mb 的 DNA 片段。

pulse labeling：脉冲标记。极短暂地提供放射性前体，使产物被放射性标记的技术。例如，将细胞置于放射性胸苷中短时间温育使 DNA 被脉冲标记。

purine：嘌呤。鸟嘌呤和腺嘌呤碱基的统称。

puromycin：嘌呤霉素。一种与氨酰 tRNA 相似的抗生素，可通过与延伸的多肽形成肽键，并从核糖体上释放出未完全合成的多肽而杀死细菌。

pyrimidine：嘧啶。胞嘧啶、胸腺嘧啶和尿嘌呤碱基的统称。

pyrimidine dimmer：嘧啶二聚体。DNA 链上通过共价连接的两个相邻的嘧啶，它们与互补链上嘌呤的碱基配对被打断。这是由紫外线照射导致的主要 DNA 损伤。

pyrrolysine：吡咯赖氨酸。“第 22 种氨基酸”。在某种古菌中通过一种特殊的 tRNA 将其加到正在延伸的多肽上。

Q

***Q*：**编码抗终止蛋白 Q 的 λ 噬菌体基因。

Q：*Q* 基因的产物。在 λ 噬菌体晚期启动子 $P_{R'}$ 下游附近的一个抑制转录终止的抗终止蛋白。

Q utilization site（*qut* site）：Q 利用位点（*qut* site）。与 λ 噬菌体晚期基因启动子重叠的 Q 结合位点。Q 结合到 *qut* 上可使 RNA 聚合酶忽视附近的终止子，并继续转录至晚期基因。

quaternary structure：四级结构。在复杂蛋白质中两个或多个多肽相互作用的方式。

quenching：猝灭。快速冷却加热变性的 DNA 使其保持变性状态。

R

R：反转录病毒中位于 U3 和 U5 区之间的 LTR 上“冗长的”区域。

R2Bm：家蚕（*Bombyx mori*）中的 LINE 类似元件。

RACE：见 rapid amplification of cDNA end。

Rad6：使 H2B 组氨酸泛素化的泛素连接酶。

RAD25：具有 DNA 解旋酶活性的酵母转录因子 TFIIH 的亚基。

radiation hybrid mapping：辐射杂种作图。将人细胞进行电离辐射处理使其染色体片段化，然后这些细胞和鼠细胞融合形成含有各种不同人染色体片段的杂合体的一种作图技术。一条染色体上紧密连锁的遗传标记趋向出现在相同的杂种细胞中。

Raf：通过信号转导蛋白 Ras 的作用而被定位于细胞膜内表面的丝氨酸/苏氨酸蛋白激酶。一旦位于细胞膜上就会被磷酸化。

Rag-1：人类 *RAG-1* 基因的产物。与 Rag-2 协同作用在 RSS 处，切割不成熟免疫球蛋白和 T 细胞受体基因，以使各种基因片段能相互重组。

Rag-2：人类 *RAG-2* 基因的产物。与 Rag-1 协同作用在 RSS 处，切割不成熟免疫球蛋白和 T 细胞受体基因，以使各种基因片段能相互重组。

***ram* state：*ram* 状态。**大肠杆菌核糖体 16S rRNA 的 H27 螺旋的不稳定状态。在此状态中，H27 螺旋的密码子和反密码子，甚至是非同类物的反密码子之间都能产生稳定的碱基配对，从而降低了解码的准确性。

RAP1：结合在特异的端粒 DNA 序列上并募集其他端粒蛋白的酵母端粒结合蛋白，包括 SIR 蛋白。

rapid amplification of cDNA end（RACE）：快速扩增 cDNA 末端（RACE）。一种扩增 cDNA 的 5′端或 3′端的方法。

rapid turnover determinant：快速周转决定子。TfR mRNA 的 3′非翻译区上的一组结构域，可确保 mRNA 较短的半衰期，但在铁离子缺乏的情况下则可提高 mRNA 的稳定性。

rare cutter：稀切酶。一种识别序列特殊，位点很少的限制性内切核酸酶。

Ras：*ras* 原癌基因的产物。在结合 CTP 的激活状态下，它可激活 Raf，由此传递信号开启刺激细胞分裂的基因。

Ras exchanger：Ras 交换因子。用 GTP 取代 Ras 上的 GDP 而激活 Ras 的蛋白。

RdRP：见 RNA-directed RNA polymerase。

reading frame：可读框。mRNA 中三联体密码的 3 种翻译方式中的一种。例如，CAGUGCUCGAC 有 3 种可能的可读框，取决于从何处开始读码：①CAG UGC UCG；②AGU GCU CGA；③GUC CUC GAC。通常 mRNA 只有一种正确的可读框。

real-time PCR：实时定量 PCR。利用荧光标记度量一段 DNA 在 PCR 中扩增量的方法。

RecA：大肠杆菌 *RecA* 基因的产物。与 SSB 一起覆盖在单链 DNA 末端，使其在同源重组中可侵入 DNA 双链寻找同源区段。也可在 SOS 应答中作为协同蛋白酶。

***recA*：**大肠杆菌中编码 RecA 蛋白的基因。

RecBCD pathway：RecBCD 途径。大肠杆菌中由 RecBCD 蛋白启动的主要的同源重组途径。

RecBCD protein：RecBCD 蛋白。在大肠杆菌同源重组过程中，在 DNA 一条链的 chi 位点附近造成缺刻从而产生一个 3′单链末端的蛋白质。

recessive：隐性的。与显性等位基因形成杂合体时不呈现表型的等位基因或性状。例如，由于 *Aa* 的表型像 *AA* 而不是 *aa*，因此 *a* 对 *A* 是隐性。

recognition helix：识别螺旋。DNA 结合蛋白质中“DNA 结合基序”的 α 螺旋，可伸入靶 DNA 的大沟并与之特异性结合，由此决定了该蛋白的特异性。实际上，识别螺旋是识别其靶 DNA 上特定的序列。

recombinant DNA：重组 DNA。两个或多个 DNA 片段间的重组产物，可在细胞中自然发生或由生物学家在体外完成。

recombination：重组。等位基因的重新组合，发生在 DNA 之间或内部的交叉行为。

recombination-activating gene：重组激活基因。编码 Rag-1 或 Rag-2 的基因。

recombination repair：重组修复。细胞用以复制含有嘧啶二聚体 DNA 的一种机制。首先，两条亲本链被复制，在二聚体处留下一个“间隙”。随后子代双链间发生重组，使“间隙”置于正常 DNA 链上，以便被填补。

recombination signal sequence（RSS）：重组信号序列。在免疫球蛋白和 T 细胞受体基因成熟过程中，被重组装置识别的位于重组交叉点上的特异序列。

recruitment：募集。促进底物聚集的一种机制。通常指募集 RNA 聚合酶或转录因子结合到启动子的事件。

regulon：调节子。受同一调节因子调控的一组基因。

related gene：相关基因。通过基因组测序计划分析发现的与相同或不同物种已知基因序列同源或部分同源的基因。

release factor（RF）：释放因子。在终止密码处导致翻译终止的蛋白质。

renaturatin of DNA：DNA 复性。见 annealing of DNA。

repetitive DNA：重复 DNA。在单倍体基因组中重复多次的 DNA 序列。

replacement vector：置换载体。源于 λ 噬菌体的克隆载体，其中很大一部分噬菌体 DNA 被去除，而

后被相似大小的外源 DNA 所取代。

replica plating：影印培养法。将菌落或细胞从一个培养皿转入另一培养皿的方法，通常用天鹅绒或其他可以从培养板上蘸取细胞的柔软材料包裹在影印铺板器上，然后将细胞印在第二个平板相对一致的位置上。

replicating fork：复制叉。两条亲本 DNA 链解链以进行复制的位点。

replicative form：复制型。单链 RNA 或 DNA 噬菌体或病毒在基因组复制过程中产生的双链形式。

replicative transposition：复制性转座。“复制并粘贴”式转座，在这一过程中转座子 DNA 被复制，一个拷贝留在原位而另一拷贝移至新位点。

replicon：复制子。从一个复制起点开始的 DNA 复制单位。

replisome：复制体。包含 DNA 复制引发体在内的多种肽链形成的大复合体，如大肠杆菌 DNA 的复制 。

reporter gene：报告基因。连接在启动子或翻译起始位点后用于检测转录或翻译活性的基因。报告基因比被替代基因更易于检测。

repressed：被阻遏。关闭，当操纵子被阻遏时，它就关闭或失活。

resolution：分离。以共联体为中间体进行转座的第二步；包括共联体分裂为两个复制子成分，每个均包含自己的转座子拷贝。也是重组的最后一步，在这一过程中链的第二次配对被断裂。

resolvase：解离酶。催化共联体解离的酶；当分支迁移后使两条 DNA 链产生缺刻从而导致 Holliday 结构解离的内切核酸酶。

res site：*res* 位点。在共联体中两个拷贝的转座子上的位点，它们之间发生交换从而完成解离。

restriction endonuclease：限制性内切核酸酶。一种识别 DNA 特异碱基序列并在此处或附近位点进行切割的酶。

restriction fragment：限制性片段。由限制性内切核酸酶将较长 DNA 片段切割形成的 DNA 片段。

restriction fragment length polymorphism（RFLP）：限制性片段长度多态性。由一种特定的限制性内切核酸酶切割两个个体的一个特定的基因位点，在切点数量上表现出的变异。

restriction map：限制性图谱。显示一段 DNA 区域内限制性位点分布的图谱。

restriction-modification system（R-M system）：限制-修饰系统。识别相同 DNA 位点的限制性内切核酸酶和 DNA 甲基酶的组合。

restriction site：限制性位点。限制性内切核酸酶识别和切割的 DNA 核酸序列。

restrictive condition：见 nonpermissive condition。

restrictive state：严谨型状态。大肠杆菌核糖体 16S rRNA 的 H27 螺旋 *ram* 状态的替代形式。在这种需要校正的状态中，H27 螺旋中要求密码子和反密码子碱基配对精确，从而使得解码的准确性增加。

retained intron：保留内含子。通过选择性剪接程序使得成熟 mRNA 中保留内含子。

retrohoming：反转录归巢。基因中的 II 类内含子转位进入基因组中某个无内含子的相同基因的过程。

retrotransposon：反转座子。一种通过类似反转录病毒的机制进行转座的转座元件，如 *copia* 或 Ty。

retrovirus：反转录病毒。一种 RNA 病毒，其复制依赖于反转录形成的原病毒。

reverse transcriptase：反转录酶。依赖于 RNA 的 DNA 聚合酶；这种通常在反转录病毒中存在的酶可催化反转录过程。

reverse transcriptase PCR（RT-PCR）：反转录 PCR（RT-PCR）。以 mRNA 为模板通过反转录酶合成 cDNA，产生的 cDNA 又作为模板进行 PCR 的过程。

reverse transcription：反转录。以 RNA 为模板合成 DNA 的过程。

reversion：回复。在同一个基因中再次发生可消除前期突变效应的突变。

RF（replicative form）：复制型。单链 DNA 噬菌体（如 ΦX174）基因组形成的环状双链形式，以便为滚环复制做准备。

RF1：原核生物中识别 UAA 和 UAG 终止密码子的释放因子。

RF2：原核生物中识别 UAA 和 UGA 终止密码子的释放因子。

RF3：原核生物中具有 GTP 酶活性的依赖核糖体的释放因子。在 GTP 存在时可帮助 RF1 和 RF2 结合到核糖体上。

RF-A：SV40 病毒的 DNA 复制所必需的人类单链 DNA 结合蛋白。

***RFN* element：**位于 *ribD* 操纵子 5′非翻译区的核开关。当与 FMN 结合后，开关中的碱基配对发生改变从而产生一个减弱转录的终止子。

RFLP：见 restriction fragment length polymorphism。

rho（ρ）：在大肠杆菌及其噬菌体的某些终止子中用以终止转录的蛋白质。

rho-dependent terminator：rho 依赖终止子。依赖 rho 起作用的终止子。

rho-independent terminator：rho 不依赖终止子。不

依赖 rho 起作用的终止子。

rho loading site：rho 装载位点。在延伸的 mRNA 链上，rho 可以结合并招募 RNA 聚合酶的位点。

ribonuclease（RNase）：核糖核酸酶。一种降解 RNA 的酶。

ribonuclease H：核糖核酸酶 H。见 RNase H。

ribonucleoside triphosphate：三磷酸核苷。RNA 的组成成分：ATP、CTP、GTP 和 UTP。

ribose：核糖。RNA 中的糖。

riboprobe：核糖核酸探针。标记的 RNA 探针，通常用于 RNA 酶保护实验中。

ribosomal RNA：核糖体 RNA。见 rRNA。

ribosome：核糖体。由 RNA 和蛋白质组成，翻译 mRNA 产生蛋白质的复合体。

ribosome recycling factor（RRF）：核糖体循环因子（RRF）。一种与 tRNA 十分相似的蛋白质，在翻译终止后结合到核糖体的 A 位点，并与 EF-G 一起作用，将核糖体从 mRNA 上释放出来。

riboswitch：核开关。RNA 上能够通过与小分子结合，进而影响转录或翻译来改变基因表达的区域。

ribozyme：核酶。具有催化功能的 RNA（RNA 酶）。

rifampicin：利福平。能抑制大肠杆菌 RNA 聚合酶起始转录的一种抗生素。

RISC loading complex（RLC）：RISC 装载复合体。分离两条 siRNA 链并将引导链送至 RISC 的蛋白质复合体。

RITS（RNA-induced initiator of transcriptional gene silencing）：RNA 诱导的转录水平基因沉默起始子。吸引 RdRP 并扩增 siRNA 的蛋白质复合体。

RLC：见 RISC loading complex。

RLIM（RING finger LIM domain-binding protein）：环指 LIM 域结合蛋白。LIM-HD 激活因子的辅阻遏物，它通过 CLIM 辅激活因子的泛素化导致蛋白质降解从而达到抑制效果。

R-looping：R 环。一种通过电镜观察 DNA 与 RNA 杂交的技术。当 RNA 取代 DNA 中的一条链与另一条链杂交时就形成典型的 R 环。R 环也可以在单链 DNA 和 RNA 间形成，但此过程不形成典型的 R 环，如果 DNA 上含有 RNA 中没有的信息，环状结构仍可观察到。

RNA（ribonucleic acid）：核糖核酸。通过磷酸二酯键相连形成的核（糖核）苷酸多聚体。

RNA-dependent DNA polymerase：依赖 RNA 的 DNA 聚合酶。见 reverse transcriptase。

RNA-directed RNA polymerase（RdRP）：RNA 指导的 RNA 聚合酶（RdRP）。一种能利用目标mRNA作为模板延伸 siRNA 引物，为 Dicer 酶和siRNA扩增提供更多底物的酶。

RNA helicase：RNA 解旋酶。一种能解开双链 RNA 或一段双链 RNA 区域的酶。

RNA-induced silencing complex（RISC）：RNA 诱导的沉默复合体（RISC）。在 RNA 干扰过程中降解靶 mRNA 的 RNA 酶复合体。包括 Dicer、Argonaute 和一种能够切割靶 mRNA 的未知 RNA 酶。

RNA interference（RNAi）：RNA 干涉。通过向细胞内插入一段双链 RNA 使特异 mRNA 降解来控制基因表达的途径。

RNA ligase：RNA 连接酶。一种能连接两段 RNA 的酶，如在前体 tRNA 中切除内含子将两个片段连接的酶。

RNA polymerase：RNA 聚合酶。一种指导转录或 RNA 合成的酶。

RNA polymerase I：RNA 聚合酶 I。真核生物中合成大的前体 rRNA 的 RNA 聚合酶。

RNA polymerase II：RNA 聚合酶 II。真核生物中合成前体 mRNA 和大多数 snRNA 的酶。

RNA polymerase III：RNA 聚合酶 III。真核生物中合成 5S rRNA 和前体 tRNA 以及几种其他前体小 RNA（包括 U6 snRNA）的酶。

RNA polymerase core：RNA 聚合酶的核心酶。具有基本延伸能力但无起始特异性转录的原核生物 RNA 聚合酶的亚基集合体；包括除 σ 因子以外的 RNA 聚合酶的所有亚基。

RNA polymerase IIA：RNA 聚合酶 IIA。带有非磷酸化或低水平磷酸化 CTD 的聚合酶 II 的一种形式。

RNA polymerase II holoenzyme：RNA 聚合酶 II 全酶。由 RNA 聚合酶 II、转录因子以及其他一些能够用温和技术纯化完整蛋白质的复合体。

RNA polymerase holoenzyme（bacterial）：RNA 聚合酶全酶（细菌）。细菌中形成全酶的多肽聚合物。通常包括 β、β′、α_2、ω 和 σ 亚基。

RNA polymerase IIO：RNA 聚合酶 IIO。CTD 高度磷酸化的一种聚合酶 II。

RNA processing：RNA 加工。通过切割、剪接、加帽、多聚腺苷酸化等修饰将初始转录产物变成成熟形式的过程。

RNase E：一种将 *E. coli* 5S rRNA 从其前体 RNA 中释放出来的酶。

RNase H：对 RNA-DNA 杂交链中的 RNA 具有特异降解作用的 RNA 酶。是反转录病毒的反转录酶的活性之一。

RNase mapping：RNase 作图。以 RNA 为探针，用 RNA 酶替代 S1 核酸酶消化单链 RNA 的衍生的 S1 作图法。

RNase P：RNA 酶 P。切除前体 tRNA 5′端的多余核苷酸的酶，大多数 RNase P 含有具催化作用的 RNA 亚基。

RNase protection assay：RNA 酶保护分析。见 RNase mapping。

RNase R：RNA 酶 R。与 tmRNA 结合并可能降解由 tmRNA 释放的非终止 mRNA 的一种核糖核酸酶。

RNase III：RNA 酶 III。在大肠杆菌前体 rRNA 加工过程中首先进行切割的酶。

RNA splicing：RNA 剪接。将内含子从转录初产物中切除并将外显子连接起来的过程。

RNA triphosphatase：RNA 三磷酸酶。在加帽之前去除前体 mRNA 5′端上的 γ 磷酸基团的酶。

rolling circle replication：滚环复制。双链环状 DNA 中的一条链保持完整并为另一条链在缺刻处的延伸作模板的 DNA 复制机制。

RRF：见 ribosome recycling factor。

rRNA（ribosomal RNA）：核糖体 RNA。定位于核糖体上的 RNA 分子。

rRNA precursor（45S）：前体核糖体 RNA（45S）。哺乳动物中的大前体 rRNA。包括 28S、18S 和 5.8S 的序列。

rRNA（5.8S）：来自哺乳动物 45S rRNA 前体的最小 rRNA，存在于核糖体大亚基（60S）中并与 28S rRNA 碱基配对。

rRNA（18S）：存在于哺乳动物核糖体小亚基（40S）中的 rRNA。

rRNA（28S）：存在于哺乳动物核糖体大亚基中与 5.8S rRNA 碱基配对的最大的 rRNA。

***rrn* gene：rrn 基因。**编码 rRNA 的细菌基因。

RSS：见 recombination signal sequence。

R2D2：推测属于 RLC 的组成部分的 Dicer 结合蛋白。

run-off transcription assay：截断转录分析。一种用于体外定量测定某基因转录程度的方法。一条含有基因调控区和 5′区的双链 DNA 在体外转录，并以标记的三磷酸核苷对产物进行标记。RNA 聚合酶不停地完全转录被截断的基因（包括末端），产生预计长度的短 RNA 产物，产物的丰度即可衡量基因在体外转录的程度。

run-on transcription assay：连缀转录分析。一种测定某基因在体内转录量的方法。首先将含有正在进行各种 RNA 链延伸反应的 RNA 聚合酶的细胞核分离出来，这些 RNA 链在体外标记核苷酸存在条件下继续延伸以标记 RNA 产物。然后将这些标记的 RNA 与含有代表被测基因的未标记样品 DNA 进行 Southern 印迹或点杂交，与印迹上每一条带或点的杂交程度是衡量来自相应基因延伸的 RNA 链的数量，由此测定这些基因转录的程度。

Ruv A：与 Ruv B 一同构成 DNA 解旋酶，在大肠杆菌同源重组过程中促使分支的迁移。

Ruv B：和 Ruv A 一同构成 DNA 解旋酶，在大肠杆菌同源重组过程中促使分支的迁移。还含有为解旋酶提供能量的 ATP 酶。

Ruv C：同源重组的 RecBCD 途径中的解离酶。

S

***Saccharomyces cerevisiae*：**酿酒酵母。

SAGA：具有乙酰转移酶活性的转录接头复合物。介导某些激活因子的效应。

SAGE：见 serial analysis of gene expression。

scanning：扫描。真核生物翻译起始模型中，40S 核糖体亚基结合到 mRNA 的 5′端并沿 mRNA 扫描或滑动，直到在合适的上下游碱基环境中发现第一个起始密码子。

scintillation：闪烁。由放射性物质发射的光脉冲在液闪计数器上激发的荧光。

scintillation fluid：见 liquid scintillation counting。

screen：筛选。一种遗传分类程序，即从含有非目标个体的群体中挑选出目的个体，但不能自动去除不需要的个体。

SC35：一种哺乳动物 SR 类 RNA 结合蛋白，可以在某些前体 mRNA 特定位点上进行定型剪接的单因子。

SDS PAGE：见 sodium dodecyl sulfate。

secondary siRNA：次级 siRNA。位于初始引发 RNA（original trigger RNA）片段以外的短的双链 RNA，并且总是位于引发 RNA（有意义链）的上游。

secondary structure：二级结构。多肽或 RNA 的局部折叠。对于 RNA 而言，二级结构可由分子内碱基配对形成。

sedimentation coefficient：沉降系数。衡量一种分子或颗粒在离心力的作用下向离心管底部运动速度的参数。

seed BAC：种子 BAC。在大基因组鸟枪法测序时，为对一个区段进行完全测序所选择的初始 BAC。

selection：选择。采用阻止生长或致死的方法淘汰非目标生物个体的遗传分类方法。

selenocysteine：硒半胱胺酸。一种稀有氨基酸，也叫做“第 21 种氨基酸”。硒取代了半胱氨酸中的硫。

SELEX（systematic evolution of ligands by exponential enrichment）：指数级富集的配体系统进化技术。

一种富集含功能区核酸（通常为 RNA）或适配体的方法。通过如亲和层析的方法将功能性分子筛选出来，随后进行 PCR 扩增，接着再反复筛选及扩增，以达到指数级富集的效果。

semiconservative replication：半保留复制。在 DNA 复制中，两条亲本链完全分离后分别与一条新的子链配对，由此形成的后代双链中均保留一条亲本链的复制方式。

semidiscontinuous replication：半不连续复制。DNA 双链分子的一条链按连续方式复制，而另一条链按不连续方式复制的模式。

sequenator：序列分析仪。DNA 自动测序仪。

sequence-tagged connector（STC）：序列标签连接子（STC）。在大规模基因组测序中，一个大克隆如 BAC 末端约 500bp 的序列。

sequence-tagged site（STS）：序列标签位点。可以通过特定引物 PCR 扩增来鉴别的短 DNA 片段。

sequencing：测序。确定一个蛋白质的氨基酸顺序或 DNA/RNA 的碱基顺序的过程。

serial analysis of gene expression（SAGE）：基因表达系列分析（SAGE）。一种可同时测定多个基因表达水平的方法。采用来自许多相关联 mRNA 对应的短 cDNA 或标签，并对其进行克隆和测序。出现频率最高的标签即为表达最活跃的。

severe combined immunodeficiency（SCID）：重症联合免疫缺陷（SCID）。在人类中无免疫系统的一种疾病。病因之一是 *Artemis* 基因的缺失，又称“气泡男孩”（bubble boy）综合征。

SH2 domain：SH2 域。在许多信号转导蛋白中发现的磷酸酪氨酸结合功能域。

SH3 domain：SH3 域。可引起蛋白质间相互作用的富含脯氨酸的螺旋结合功能域。

Shine-Dalgarno（SD）sequence：SD 序列。一段与大肠杆菌 16S rRNA 3′端序列互补的富含 AG 的序列（多数为 AGGAGGU）。这两段序列间的碱基配对可协助核糖体与 mRNA 结合。

short hairpin RNA（shRNA）：短发夹 RNA（shRNA）。构建的具有反向重复的 RNA 以使其在体内形成发夹结构并启动 RNA 干扰。

short interfering RNA：见 siRNA。

shotgun sequencing：鸟枪测序法。一种基因测序法。先将基因组切成很多小片段并克隆后进行随机测序，然后将这些小片段连接起来以获得整个基因组的序列。

shuttle vector：穿梭载体。一种能够在两个或多个不同宿主中复制的克隆载体，允许重组 DNA 在不同宿主间穿梭。

sickle cell disease：镰状细胞病。一种产生异常 β-球蛋白的遗传疾病。由于一个氨基酸的改变使得血红蛋白在低氧情况下产生聚集，从而使血红细胞扭曲成镰刀形。

sigma（σ）：原核生物 RNA 聚合酶中赋予转录特异性的一个亚基，具有识别特异启动子的能力。

signal peptide：信号肽。一段约 20 个氨基酸且通常位于多肽的氨基端的序列，可帮助新合成的多肽及其核糖体锚定于内质网。带有信号肽的多肽将定向转运到高尔基体中包装并运出细胞。

signal transduction pathway：信号转导途径。一种连接信号（如生长因子结合在细胞表面）与细胞内基因激活（或抑制）效应的生化途径。

silencer：沉默子。可在远距离降低真核基因转录的 DNA 元件。

silencing：沉默。真核基因活性的抑制，可通过在基因区形成异染色质或更小范围内的某种机制包括特定核小体与 DNA 间更紧密的结合来实现。

silent mutation：沉默突变。在有机体中产生的一种即使在单倍体或纯合子的情况下也不可察觉的突变。

SIN3：一种酵母的辅抑制因子。

SIN3A and SIN3B：哺乳动物中与酵母 SIN3 对应的辅抑制因子。

single-nucleotide polymorphism（SNP）：单核苷酸多态性（SNP）。某个特异基因位点在两个或多个个体间的单个核苷酸差异。

single-strand DNA-binding protein：见 SSB。

siRISC：果蝇 RNA 干扰中有 siRNA 参与的 RISC。

SIR2、SIR3 and SIR4：酵母异染色质（包括端粒异染色质）形成所需的相关蛋白质。

siRNA（short interfering RNA）：短的干扰 RNA。双链引发 RNA 的短片段（21～28nt）由 RNA 过程中 Dicer 酶作用产生。

site-directed mutagenesis：定点诱变。将预期的特异变异引入到一个克隆基因中的方法。

site-specific recombination：特定位点重组。依赖于有限的序列相似性，常发生在相同位置的重组。

Ski complex：Ski 复合物。能够降解真核生物中非终止 mRNA，包括 Ski7p 和外切核酸酶在内的复合体的一部分。

Ski7p：识别真核生物核糖体 A 位点上 mRNA 末端 poly（A）的非终止 mRNA 的蛋白质，它能聚集 RNA 酶复合体降解非终止 mRNA。

SL1：包含 TBP 与 3 种 TAF_I 的 I 型转录因子。与 UBF 协同激活聚合酶 I 结合 DNA 并转录。

slicer：RNA 切割酶。在 RISC 中切割靶 mRNA 的具

有 Argonaute2 活性的酶。

sliding clamp：滑动钳。RNA 聚合酶 II 中位于臂末端和底架之间的钳状结构，保证了酶与 DNA 模板的分离，因而增强了前进的连续性。

Slu7：在内含子 3′剪切末端选择合适的 AG 进行剪接所必需的因子。

small nuclear RNA（snRNA）：核内小 RNA（snRNA）。在细胞核中发现的一系列小 RNA，与蛋白质结合形成核内小核糖核蛋白体（snRNP），参与前体 mRNA 的剪接。

small nucleolar RNA（snoRNA）：核仁小 RNA（snoRNA）。在核仁中发现的一系列大量存在的小 RNA。其中 snoRNA 的一小部分（E1、E2、E3），与蛋白质结合形成核仁小核糖核蛋白（snoRNP），参与前体 rRNA 的加工。

small RNA（sRNA）：小 RNA。细菌中一类短的 RNA 分子，与 mRNA 结合可控制翻译。

SMCC/TRAP：人类调节子蛋白质。

Sm protein：Sm 蛋白。在所有 snRNP（包括次要的 snRNP）中发现的一组七蛋白体。

Sm site：Sm 位点。在 snRNA 上与 SM 蛋白相互作用的一段序列（AAUUUGUGG）。

SmpB：与 tmRNA 结合并帮助其结合到核糖体上的蛋白质，可能在某种程度上弥补了 tmRNA 上 D 环的缺失。

snoRNA：见 small nucleolar RNA。

snoRNP：见 small nucleolar RNA。

SNP：见 single-nucleotide polymorphism。

snRNA：见 small nuclear RNA。

snRNP：见 small nuclear RNA。

sodium dodecyl sulfate（SDS）：十二烷基磺酸钠。一种带强负电荷的洗涤剂，用于 SDS 聚丙烯酰胺凝胶电泳（SDS-PAGE）中蛋白质的变性。

somatic cell：体细胞。

somatic 5S rRNA gene：体细胞 5S rRNA 基因。在体细胞和卵母细胞中均表达的 5S rRNA 基因（在非洲爪蟾单倍体中的数目大约是 400 个）。

somatic mutation：体细胞突变。一种只影响体细胞的突变，因此不会传给下一代。

S1 mapping：S1 核酸酶定位法。绘制特异转录产物末端图谱（或对转录产物定量）一种方法。被标记的 DNA 探针与体内或体外合成的转录物杂交，用 S1 核酸酶去除杂交产物中未杂交部分，并将被探针保护的部分进行电泳来确定其大小。

S1 nuclease：S1 核酸酶。单链 RNA 和 DNA 的一种特异核酸酶。用于 S1 核酸酶定位。

Sos：一种 Ras 交换子。

SOS response：SOS 应答。为协助大肠杆菌细胞对环境胁迫（例如化学诱变和辐射等）作出响应而进行的包括 recA 在内的一系列基因的激活。

Southern blotting：Southern 印迹法。将凝胶电泳分离的 DNA 片段转移至合适的支持介质（如硝酸纤维素膜）上，以备与标记探针杂交。

spacer DNA：间隔 DNA。位于重复基因（如 rRNA 基因）之间或内部的 DNA 序列。

specialized gene：特化基因。仅在某一（或极少）类型细胞中具有活性的基因。例如，β 胰岛细胞中的胰岛素基因，血红细胞中的球蛋白基因。又称为奢侈基因。

spliced leader（SL）：剪接前导序列。锥形虫中独立合成的 35 核苷酸的前导序列，被反式剪接至表面抗原基因 mRNA 的编码区内。

spliceosome：剪接体。大的 RNA-蛋白质复合体，是核内前体 mRNA 进行剪接的场所。

spliceosome cycle：剪接体循环。剪接体形成、剪接及解离的过程。

splicing：剪接。将 RNA 的内含子去除并将两段外显子连接起来的过程。

splicing factor：剪接因子。包括 snRNP 在内的、对核前体 mRNA 剪接起关键作用的蛋白质。

Spo11：酵母中能够产生 DSB 并启动减数分裂重组的内切核酸酶。

spore：孢子。①由植物或真菌经有性过程，或真菌经无性过程形成的特异单倍体细胞。后者也可作为配子，或萌发产生新的单倍体细胞。②某种细菌在逆境下经无性过程产生的一种特异细胞。这种孢子惰性相对较强且对逆境具抗性。

sporulation：孢子发生。

squelching：压制。通过增加一个激活因子的浓度来抑制另一个激活因子。可能是由对稀缺的通用因子的竞争而引起。

SRB and MED-containing cofactor（SMCC）：见 SMCC/TRAP。

SRC：见 steroid receptor coactivator。

SRC-1，SRC-2 and SRC-3：见 steroid receptor coactivator（SRC）。

sRNA：见 small RNA。

SR protein：SR 蛋白质。一组富含丝氨酸和精氨酸与 RNA 结合的蛋白质。

SSB：单链 DNA 结合蛋白，参与 DNA 复制和重组过程。与单链 DNA 结合以阻止其与互补链的碱基配对。

STC：见 sequence-tagged connector。

steroid receptor coactivator（SRC）：类固醇受体辅激

活因子（SRC）。一个结合带有配体的类固醇受体的蛋白家族成员，能帮助这些激活因子富集 CBP。

sticky end：黏性末端。能通过互补和碱基配对连接的双链 DNA 的单链末端。

Stn1p：见 Cdc13p。

stop codon：终止密码子。编码翻译终止的 3 种密码子（UAG、UAA、UGA）。

stopped-flow apparatus：停流装置。一种进行动力学实验的装置，各种试剂在其中被快速地聚集在一起，并能检测其反应过程。

strand exchange：见 postsynapsis。

streptavidin：链亲合素。一种由链霉菌产生的易与生物素结合的蛋白质。

streptomycin：链霉素。一种使核糖体误读mRNA而杀死细菌的抗生素。

stringency（of hybridization）：（杂交的）严谨度。影响两条多核苷酸链杂交能力的所有因素（温度、盐离子、有机溶剂的浓度等）的组合。在高度严谨条件下，只有完全互补的链才会杂交，在低严谨条件下，可以产生一些错配。

strong-stop DNA：强终止 DNA。反转录病毒 RNA 的一种反转录初始产物。它起始于引物结合位点，终止在病毒 RNA 5′端下游 150nt 处。

structural genomics：结构基因组学。对基因组序列的研究（原文为：The study of expression）。

structural variation（genomic）：基因组结构变异。大片段 DNA 的变异，与 SNP 相反。

STS：见 sequence-tagged site。

SUMO（small ubiquitin-related modifier）：小泛素相关修饰因子。能够附加在其他蛋白质（如激活因子）上的小肽，可以将蛋白质靶定到细胞核内的一个隔离区内，使激活因子不能激活基因的表达。

sumoylation：SUMO 化。将 SUMO 蛋白附加在其他蛋白质上。

supercoil：见 superhelix。

superhelix：超螺旋。环状双链 DNA 的一种形式，其中双螺旋像搓捻的橡胶带自我缠绕。

superinfection（λ phage）：重复感染。一个 λ 噬菌体的溶源细菌又被另一个 λ 噬菌体感染。

supershift：超变动。当新的蛋白质结合到蛋白质-DNA 复合体时所引起的凝胶电泳迁移率额外变动的现象。

suppression：抑制。一种突变对另一种突变的补偿。

suppressor mutation：抑制子突变。能够逆转相同或不同基因上突变效果的突变。

SV40：猿猴病毒 40。具有小环状基因组的 DNA 瘤病毒，能对某些啮齿动物致瘤。

SWI/SNF：通过破坏核小体核心帮助染色质重建的辅激活因子家族。

synapsis：联会。在大肠杆菌同源重组过程中，参与交换的单链和双链 DNA 中互补序列的并行排列。

syntenic block：同线群。在不同生物间具有保守基因顺序的 DNA 群。

synteny：同线性。不同生物间基因顺序的保守性，即在给定生物中相同染色体上基因的排列方式。

synthetic lethal screen：合成致死筛选。对相互作用基因的一种筛选方法。利用细胞带有非致死性突变（即条件致死突变）的一种基因搜寻其他基因，在这些基因中通常的非致死性突变却是细胞致死性的。这可能意味着，这两种基因的产物以某种方式相互作用，其中任意一种有缺陷是不会致死的，但两种都有缺陷则是致死的。

T

TAF：见 TBP-associated factor。

tag（in SAGE）：标签（在 SAGE 中）。细胞中对应于 mRNA 的短 cDNA 片段。

tag SNP：标签 SNP。揭示同一区段上其他 SNP 特点的判断性 SNP。

T antigen：T 抗原。DNA 瘤病毒 SV40 的早期基因区域的主要产物。是具有 DNA 解旋酶活性的 DNA 结合蛋白，能够转化细胞从而致瘤。

***Taq* polymerase：*Taq* 聚合酶。**从水生嗜热杆菌（*Thermus aquaticus*）中分离出来的耐高温的 DNA 聚合酶。

target mRNA：靶 mRNA。在 RNA 干扰中被靶向，降解的 mRNA。

TATA box：TATA 盒。在大多数被 RNA 聚合酶 II 识别的真核生物启动子中，位于转录起始位点上游 25～30 个碱基处，具有 TATAAAA 保守序列的元件称为 TATA 盒。

TATA-box-binding protein（TBP）：TATA 盒结合蛋白（TBP）。分别在 I、II、III 类前起始复合物中的 SL1、TF II D 及 TF III B 因子中的亚基。与含有 TATA 盒的 II 类启动子的 TATA 盒结合。

t loop：t 环。在真核生物染色体的端粒中形成的环。

T loop：T 环。tRNA 分子中具有几乎不变的 TψC 序列（ψ 指假尿苷）的环，通常画在结构图的右侧。

TBP：见 TATA-box-binding protein。

TBP associated factor：TBP 结合因子。在 SL1、TF_{II} D或 TF_{III} B 因子中与 TBP 结合的蛋白质。

TBP-free TAF II-containing complex（TFTC）：缺乏 TBP 的可替换的 TF_{II} D 因子。

TBP-like factor（TLF）：类 TBP 因子（TLF）。TBP

的同源物，缺少能插入苯丙氨酸以帮助 TATA 盒弯曲的功能。作为 TBP 的替代物至少可与某些缺乏 TATA 盒的 II 型启动子结合。

TBP-related factor 1（TRF1）：TBP-相关因子 1（TRF1）。在果蝇神经发育中发现的具有活性的另一种 TBP。

TC-NER：见 transcription-coupled NER。

T-cell receptor（TCR）：T 细胞受体。T 细胞表面的抗原结合蛋白。由两条重链（β）和两条轻链（α）组成。

T-DNA：Ti 质粒上诱导肿瘤发生的一段 DNA。

TEBP：见 telomere end-binding protein。

telomerase：端粒酶。一种能够在 DNA 复制之后延伸端粒末端的酶。

telomere：端粒。真核生物中含有短 DNA 序列串联重复结构的染色体末端。

telomere end-binding protein（TEBP）：端粒末端结合蛋白（TEBP）。存在于纤毛原生动物中，与端粒末端结合使其不被降解和 DNA 修复酶修饰的一种二聚体蛋白。

telomerase reverse transcriptase（TERT）：端粒酶反转录酶（TERT）。具有反转录酶活性位点的端粒酶亚基。

telomere position effect（TPE）：端粒位置效应。沉默近端粒基因的效应。

temperate phage：温和噬菌体。能够进入溶源期并形成原噬菌体的噬菌体。

temperature-sensitive mutation：温度敏感型突变。一种在高温（非许可温度）下产生有缺陷产物，而低温（许可温度）下产生有功能产物的突变。

template：模板。指导合成互补多聚核苷酸的一条多核苷酸链（RNA 或 DNA），如一条 DNA 链可作为转录的模板。

template DNA strand：模板 DNA 链。基因中与 RNA 产物互补的 DNA 链，即合成 RNA 的模板链。有时又称反密码链或反义链。

Ten1p：见 Cdc13p。

teratogen：致畸剂。导致生物异常发育的物质。

terminal transferase：末端转移酶。一种将脱氧核糖核酸一次一个地加至 DNA 3′端的酶。

terminal uridylyl transferase（TUTase）：末端尿苷酰转移酶。在 RNA 编辑过程中把 UMP 残基加至前体 mRNA 上的酶。

terminator：终止子。见 transcription terminator。

***Ter* site：*Ter* 位点。**在大肠杆菌 DNA 复制的终止区域的 6 个位点，分别为：*TerA*、*TerB*、*TerC*、*TerD*、*TerE* 和 *TerF*。

TERT：见 telomerase reverse transcriptase。

tertiary structure：三级结构。多肽或 RNA 的全方位的三维立体结构。

tetramer（protein）：四聚体。4 个多肽形成的复合体。

TF II A：能够使 TF II D 与 TATA 盒稳定结合的 II 类通用转录因子。

TF II B：在体外于 TF II D 因子之后结合到启动子上的 II 类通用转录因子，帮助 TF II F 和 RNA 聚合酶 II 结合到启动子上。

TF II B_C：TF II B 因子中与 TBP 致弯曲后的 TATA 盒结合的 C 端功能域。

TF II B_N：TF II B 因子中结合于 RNA 聚合酶 II 活性位点附近的 N 端功能域，使 RNA 聚合酶 II 定位于距 TATA 盒正确的位置以便起始转录。

TF II D：在体外首先结合到含 TATA 盒的启动子上的 II 类通用转录因子，并作为前起始复合物组装的集结位点。包含 TATA 盒结合蛋白（TBP）和 TBP 结合因子（$TAF_{II}s$）。

TF II E：在体外与 TF II F 和 RNA 聚合酶之后，但在 TF II H 之前结合到前起始复合体上的 II 类通用转录因子。

TF II F：在体外与 TF II B 结合之后并与 RNA 聚合酶 II 协同结合到前起始复合体上的 II 类通用转录因子。

TF II H：在体外最后结合到前起始复合体上的 II 类通用转录因子。具有蛋白激酶和 DNA 解旋酶活性。

TF II S：通过限制 RNA 聚合酶 II 在暂停位点的停留，激发转录延伸的蛋白质。

TF III A：通过促进 TF III B 的结合，与 TF III C 共同协助激活真核生物 5S rRNA 基因转录的一种通用转录因子。

TF III B：通过结合基因的上游区域来激活由 RNA 聚合酶 III 调控基因转录的通用转录因子。

TF III C：激发 TF III B 与经典的 III 型基因结合的通用转录因子。

TfR：见 transferrin receptor。

TFTC：见 TBP-free TAFII-containing complex。

thermal cycler：热循环仪。通过在引物退火、DNA 延伸和 DNA 变性 3 种温度之间不断循环而自动进行 PCR 反应的仪器。

thymidine：胸腺嘧啶脱氧核苷。含有胸腺嘧啶的核苷酸。

thymine（T）：胸腺嘧啶。在 DNA 中与腺嘌呤配对的嘧啶碱基。

thymine dimer：胸腺嘧啶二聚体。在同一条 DNA 链

中两个相邻的胸腺嘧啶共价连接在一起，干扰与其互补链上对应腺嘌呤间的配对。

thyroid hormone receptor (TR)：甲状腺激素受体。一种细胞核受体。当缺乏甲状腺素时起抑制作用，而甲状腺素存在时则成为激活因子。

thyroid hormone receptor associated protein (TRAP)：见 SMCC/TRAP。

thyroid hormone response element (TRE)：甲状腺激素响应元件。对甲状腺素受体与甲状腺素复合物响应的增强子。

TIF-1B：存在于某些低等真核生物中，与人类的 SL1 同源。

Ti plasmid：Ti 质粒。来自根瘤农杆菌中的致瘤质粒。通常用作携带外源基因进入植物细胞的载体。

TLF：见 TBP-like factor。

tmRNA：见 transfer-messenger RNA。

tmRNA-mediated ribosome rescue：tmRNA 介导的核糖体拯救。用 tmRNA 在缺乏终止密码子的 mRNA上拯救停运的核糖体。tmRNA 负载丙氨酸后进入停运核糖体的 A 位点，并将丙氨酸供给停止合成的多肽。在转位过程中，这个具有短可读框的 tmRNA 转至 P 位点，取代非终止的 mRNA。

Tn3：大肠杆菌中编码氨苄青霉素抗性基因的转座子。

toeprint assay：趾纹分析。测定与 DNA 或 RNA 结合蛋白质的边界所用的引物延伸分析方法。

topo IV：拓扑异构酶 IV 。在大肠杆菌 DNA 复制结束时使子代双螺旋脱连环的酶。

topoisomerase：拓扑异构酶。改变 DNA 超螺旋或拓扑结构的酶。

type I topoisomerase：I 型拓扑异构酶。将瞬时单链缺刻引入 DNA 底物的酶。

type II topoisomerase：II 型拓扑异构酶。将瞬时双链缺刻引入 DNA 底物的酶。

torus：一种圆环结构。

TPA：12-*O*-十四烷酰佛波醇-13-醋酸盐。一种佛波酯。

TPE：见 telomere position effect。

trailer：尾随序列。位于 mRNA 3′端的终止密码子与 poly（A）序列之间的非翻译碱基区域，也称 3′-UTR。

***trans*-acting：反式作用。**用于描述基因元件的专有名词，如阻遏物基因或转录因子基因，它们可以影响位于不同染色体上的其他基因。这些反式作用基因通过产生能在远距离作用的可扩散物质发挥功能。

transcribed fragment (transfrag)：转录片段。在微阵列检测中，由转录物所代表的基因组的一小段区域。

transcribed spacer：(可) 转录间隔区。前体 rRNA 中的一部分区域，在形成成熟 rRNA 的加工过程中被去除。

transcript：转录物。一个基因的 RNA 拷贝。

transcription：转录。从基因到其 RNA 拷贝的合成过程。

transcription-activating domain：转录激活域。转录激活因子中激活转录的部分。

transcription arrest：转录停滞。在转录过程有时会发生永久暂停，这类停滞需要借助外力才能重新开始转录。

transcription bubble：转录泡。在合成 RNA 过程中紧随 RNA 聚合酶之后的局部解链的 DNA 区域。

transcription-coupled NER (TC-NER)：转录偶联的核苷切除修复。一种核苷酸切除修复，因为需要 RNA 聚合酶检测损伤并召集 NER 装置，因此只能去除转录链的损伤。

transcription factor：转录因子。通过结合启动子或增强子元件激发（或有时抑制）真核基因转录的蛋白质。

transcription pause：转录暂停。转录的临时中断，通常能被聚合酶所逆转。

transcription terminator：转录终止子。引发转录终止的一段特异的 DNA 序列。

transcription unit：转录单位。由启动子和终止子界定的一段 DNA 区域，也是一个独立的单位。可能含有多个编码区，如在腺病毒中的主要晚期转录单位。

transcriptional mapping：转录作图。将转录物（并非只是基因）定位于基因组特定位点上。

transcriptome：转录组。某一生物一生中所产生的所有转录物的总和。

transcriptomics：转录组学。对某种生物的转录物进行全面研究的科学。

transcript of unknown function (TUF)：未知功能的转录物 (TUF)。功能未知的非编码蛋白的转录物。

transesterification：转酯（反应）。打开一个酯键并同时生成另一个酯键的反应，如细胞核前体 mRNA 剪接索套中间体的形成就是转酯反应的过程。

transfection：转染。将外源 DNA 转入真核生物的细胞。

transferrin：运铁蛋白。通过与运铁蛋白受体结合将铁输入细胞内的一种含铁蛋白质。

transferrin receptor (TfR)：运铁蛋白受体。结合运

铁蛋白且允许与其有效负载的铁进入细胞的一种膜蛋白。

transfer-messenger RNA（tmRNA）：转移信使 RNA（tmRNA）。类似 tRNA 且能拯救非终止 mRNA 上停运核糖体的一段 300nt 的 RNA。

transfer RNA：见 tRNA。

transformation（genetic）：遗传转化。导入外源 DNA 使细胞基因组成发生变化的过程。

transgene：转基因。将外源基因移植到某种生物体内并使其成为转基因生物的过程。

transgenic organism：转基因生物。转入一个或一组新基因的生物。

transition：（碱基）转换。由嘧啶取代嘧啶或嘌呤取代嘌呤所导致的突变。

translation：翻译。核糖体利用 mRNA 携带的信息合成蛋白质的过程。

translesion synthesis（TLS）：（DNA）跨损伤合成。通过复制绕过 DNA 损伤区域的修复机制。

translocation：转位。翻译延伸阶段中紧接肽酰转移酶反应之后的步骤，此过程中 mRNA 在核糖体上移动一个密码子长度并让一个新的密码子进入核糖体 A 位点。

transmission genetics：传递遗传学。研究基因从一代向下一代传递的科学。

transpeptidation：转肽作用。由肽酰转移酶催化的（形成肽键的）反应。

transposable element：转座元件。可以从基因组的一个位点移动至另一位点的 DNA 元件。

transposase：转座酶。由转座子编码并催化转座的一类蛋白质的总称。

transposition：转座。DNA 元件（转座子）从一个位点到另一个位点移动的过程。

transposon：见 transposable element。

***trans*-splicing：反式剪接。**将两个不同转录单位合成的 RNA 片段剪接在一起。

***trans*-translation：反式翻译。**发生在 tmRNA 介导的核糖体拯救过程中，由阅读非终止 mRNA 到阅读 tmRNA 可读框之间的转换。

transversion：颠换。由嘧啶替换嘌呤或嘌呤替换嘧啶所导致的突变。

***trc* promoter：*trc* 启动子。**用于多种表达载体的启动子，它包含 trp 启动子的－35 框，以提供强度，lac 启动子的－10 框和操纵基因，以提供诱导性。

TRE：见 thyroid hormone response element。

TRF1 and TRF2（TTAGGG repeat-binding factor）：TTAGGG 重复序列结合因子。特异地结合端粒的双链 DNA 的端粒结合蛋白。

TRF1：见 TBP-related factor 1。

trigger dsRNA：触发双链 RNA。引发 RNA 干涉的双链 RNA。

Trl：一种 GAGA 框结合蛋白。

tRNA（transfer RNA）：转运 RNA。一端结合氨基酸，另一端识别 mRNA 密码子的小分子 RNA，充当由 mRNA 密码翻译成氨基酸序列过程中的适配器作用。

tRNA charging：tRNA 负载。由氨酰-tRNA 合成酶催化，将 tRNA 与相应氨基酸相偶联的过程。

$tRNA_i^{Met}$：细菌中负责起始蛋白质合成的 tRNA。

$tRNA_i^{Met}$：真核生物中负责起始蛋白质合成的tRNA，其类似物有 $tRNA_i^{Met}$。

$tRNA_f^{Met}$：将甲硫氨酸插入肽链内部的 tRNA。

tRNA A-like domain：tRNA A 类似结构域。一种 5′端和 3′端碱基配对的 tmRNA，其类似于 tRNA 的接受臂。

***trp* operon：色氨酸操纵子。**编码合成色氨酸的一系列酶的操纵子。

***trp* repressor：色氨酸阻遏物。**色氨酸操纵子的阻遏物，由色氨酸阻遏物蛋白和辅阻遏物色氨酸共同组成。

trypanosome：锥形虫。寄生在哺乳动物和采采蝇体内的原生动物；后者通过叮咬哺乳动物传播疾病。

T segment：T 节。一种可在拓扑异构酶 II 作用下穿过 G 片段通道的 DNA 片段。

tumor suppressor gene：肿瘤抑制基因。可控制细胞分化从而抑制恶性肿瘤发展的基因。

Tus：*Ter* 利用底物。结合 *Ter* 位点且参与复制终止的一种大肠杆菌蛋白。

TUTase：见 terminal uridylyl transferase。

two-dimensional gel electrophoresis：双向凝胶电泳。一种高分辨率分离蛋白质的方法。首先通过等电聚焦进行蛋白质的一维分离，然后经 SDS-PAGE 进行二维分离。

Ty：通过类似于反转录病毒机制进行转座的酵母转座子。

U

U1 snRNP：第一个识别细胞核前体 mRNA 5′端剪接位点的 snRNA。

U2AF35 and U2AF65：U2AF 的两个亚基。

U2-associated factor（U2AF）：U2 结合因子。通过结合 3′剪接位点的多聚嘧啶区和 AG 序列，帮助识别 3′剪接位点上正确 AG 的剪接因子。

U2 snRNP：识别细胞核前体 mRNA 分支位点的 snRNA。

U3：反转录病毒中的长末端重复序列（LTR）上的3′端非翻译区。

U4 snRNP：snRNP的一种组分，在U6 snRNA开始对核前体mRNA进行剪接之前，它的RNA与U6 snRNP中的RNA实现碱基配对。

U4atac：具有与U4 snRNA同样功能的一种较小的snRNA。参与对变异的内含子的剪切过程。

U5：反转录病毒的LTR上的5′端非翻译区。

U5 snRNP：同时与外显子-内含子接合区的5′端和3′端结合从而在剪接中帮助两个外显子相互靠近的一种snRNP。

U6 snRNP：snRNP的一种组分，在剪接体中其RNA与前体mRNA中的5′端剪接位点及U2 snRNP中的RNA实现碱基配对。

U6atac：具有与U6 snRNA同样功能的一种较小的snRNA。参与对变异的内含子的剪切过程。

U11 snRNP：具有与U1 snRNA同样功能的一种较小的snRNA。参与对变异的内含子的剪切过程。

U12 snRNP：具有与U2 snRNA同样功能的一种较小的snRNA。参与对变异的内含子的剪切过程。

UAF：见upstream activating factor。

ubiquitin：泛素。一种能单个或成串地附加在蛋白质（包括激活因子）上的小多肽。单一泛素化通常对激活因子具有激活效应，而多聚泛素化标记的蛋白质将运至蛋白酶体降解。

ubiquitinated protein：泛素化蛋白。至少附加了一个泛素分子的蛋白质。

UAS_G：见upstream activating sequence。

UBF：见upstream-binding factor。

ultraviolet（UV）radiation：紫外辐射。来自日光中的紫外线辐射可导致DNA产生嘧啶二聚体。

UmuC：$UmuD'_2C$复合体组分的一种。

umuC：*umuDC*操纵子上的一种基因，由DNA损伤SOS应答诱导。

UmuD：$UmuD'_2C$复合物的一种组分，可由蛋白酶修剪形成UmuD′。

umuD：一种*umuDC*操纵子基因，被DNA损害引起的SOS响应所诱导。

***umu* DC**：含有*umuD*和*umuC*基因的操纵子。

$UmuD'_2C$：也被称作DNA聚合酶V，能忽略由嘧啶二聚体引起的易错突变。

undermethylated region：低甲基化区。基因内或其两侧的相对较少或无甲基的一段区域。

unidirectional DNA replication：单向DNA复制。沿着一个方向且仅有一个活跃复制叉的一种复制。

unit cell：晶胞。晶体中的小重复单元。

untranslated region（UTR）：非翻译区。位于mRNA编码区外的5′端或3′端不被翻译的区域。

UP element：上游元件。在某些细菌强启动子的−35框上游发现的额外启动子元件，能额外增强聚合酶与启动子间的相互作用。

Upf1 and Upf2：哺乳动物T细胞下游不稳定元件的一部分。

up mutation：上升突变。通常发生在启动子中并导致基因表达增强的一种突变。

UPE：见upstream promoter element。

upstream activating sequence（UAS_G）：上游激活序列。酵母中结合于激活因子GAL4上并作用于半乳糖利用基因的增强子。

upstream activating factor（UAF）：上游激活因子。UBF的酵母同源物。

upstream-binding factor（UBF）：上游结合因子。结合在上游启动子元件（UPE）上的I型转录因子，与SL1协同促进聚合酶I的结合和转录。

upstream-binding site（UBS）：上游结合位点。结合终止子上RNA发夹结构的上半部分，并由此延缓发夹结构的形成，随后抑制转录终止的细菌核心RNA聚合酶上的位点。

upstream promoter element（UPE）：上游启动子元件。在某些真核生物I型和II型启动子中发现的位于核心启动子上游的启动子元件。这两类启动子中的上游元件序列是完全不同的。

uracil（U）：尿嘧啶。在RNA中与胸腺嘧啶类似的一种嘧啶碱基。

uracil *N*-glycosylase：尿嘧啶*N*-糖基化酶。从DNA链中除去尿嘧啶从而留下一个无碱基（无碱基的糖）位点的酶。

uridine：尿苷。含尿嘧啶的核苷酸。

UTR：非翻译区。在mRNA的3′端或5′端不被翻译的区域。

V

variable loop：可变环。位于tRNA分子中的反密码子和T环之间的环或茎。

variable number tandem repeat（VNTR）：可变串联重复数。在限制性内切核酸酶切割位点间含有串联重复的微卫星序列的一类RFLP。

variable region：易变区域。抗体上特异结合外源物质或抗原的区域。顾名思义，此区域在不同抗体间有很大变化。

V（D）J joining：V（D）J连接。包括分离胚胎基因的V和J或V、D和J等片段及随后的重组事件在内，形成有活性免疫球蛋白或T细胞受体基因的组装过程。

vector：载体。在基因克隆实验中作为携带者的DNA（质粒或噬菌体DNA）。

vegetative cell：营养细胞。仅通过分裂而不是形成孢子或有性繁殖进行增殖的细胞。

virulent phage：烈性噬菌体。裂解宿主的噬菌体。

void volume：外水体积。凝胶层析实验中，包含大分子不能进入胶孔的部分。

VP16：带有酸性转录激活域（但没有DNA结合域）的疱疹病毒转录因子。

W

Western blot：Western印迹法。蛋白质经电泳并转膜，然后与特异抗体或抗血清反应，最后抗体由标记的二抗或蛋白A进行检测。

wobble：摇摆。密码子的第三个碱基有可能与反密码子的第一个碱基形成非沃森-克里克的碱基配对方式，从而允许tRNA解读更多密码子。

wobble base pair：摇摆碱基配对。通过摇摆原理（如G-U或A-I的配对）形成的碱基配对。

wobble hypothesis：摇摆假说。Francis Crick用摇摆碱基配对来解释一个反密码子可解读多个密码子的假说。

wobble position：摇摆位点。摇摆碱基配对原则允许的密码子的第三个碱基。

wyosine：Y核苷。tRNA中高度修饰的一种鸟嘌呤核苷酸。

X

xeroderma pigmentosum（XP）：着色性干皮病。对阳光极度敏感的疾病，甚至暴露于温和的阳光下也会导致许多皮肤癌变，是由核苷切补修复缺陷所导致。

Xis：λ*xis*基因产物，负责宿主DNA中对λDNA的切除。

XPA：证实损伤DNA是否已经与XPC结合并帮助人类NER复合物的其他成分进行组装的一种蛋白质。

***XPA-XPG*：**与核苷酸切除修复（NER）相关的人类基因。任何一个基因的突变都会导致着色性干皮病（XP）。

***XPB*：**人类TF II H DNA解旋酶的两个亚基之一，在NER过程中的DNA的解链所必需。

***XPC*：**与另一种蛋白质一起识别DNA损伤并起始人类GG-NER。

***XPD*：**人类TF II H DNA解旋酶的两个亚基之一。在NER过程中对于DNA的解链是必需的。

***XPF*：**和ERCC1一起在人类NER过程中对DNA损伤区的5′端进行切割。

***XPG*：**一种在人类NER过程中对DNA损伤区的3′端进行切割的内切核酸酶。

XP-V：DNA聚合酶η基因突变所导致的一种着色性干皮病（XP）的变异形式。

X-ray crystallography：X射线晶体学。见X-ray diffraction analysis。

X-ray diffraction analysis：X射线衍射分析。一种通过晶体分子的X射线衍射来确定分子三维结构的方法。

X-ray：X射线。晶体衍射产生的高能量辐射。X射线衍射谱可用于确定晶体的分子形状，X射线还可使细胞成分离子化从而造成染色体的破坏。

Xrn2：一种人类5′→3′外切核酸酶，降解共转录切割后的下游RNA产物，引起转录终止。

Y

YAC：见yeast artificial chromosome。

yeast artificial chromosome：酵母人工染色体。一种高容量的克隆载体，由酵母的左右两个端粒和着丝粒组成。置于着丝粒和一个端粒之间的DNA成为YAC的一部分并在酵母细胞中被复制。

yeast two-hybrid assay：酵母双杂交检测技术。分析两种蛋白质互作的技术。其中一种蛋白质（称作“饵”）与来自另一蛋白质的DNA结合域形成融合蛋白。其他蛋白质（靶）与转录激活域形成融合蛋白。如果两种融合蛋白在细胞中相互作用，它们就形成能够激活一个或多个报告基因的激活因子。

Z

Z-DNA：双链DNA左手螺旋模式，其骨架呈现之字形结构。这种结构往往通过嘌呤和嘧啶间的交互伸展作用而稳定。

zinc finger：锌指结构。一种与DNA结合的模体，其结构由1个锌离子和4个氨基酸组成，通常是2个半胱氨酸和2个组氨酸。此模体呈手指形状，与DNA的大沟特异结合。

α-amanitin：α-鹅膏（蕈）毒环肽。在鹅膏菌属的多种有毒蘑菇中产生的一种毒素。当毒素的浓度很低时，可抑制RNA聚合酶II的活性，而当浓度提高时，则会抑制RNA聚合酶III的活性。通常不抑制RNA聚合酶I的活性。

β clamp：β钳。DNA聚合酶III全酶中β亚基的二聚体，它能夹住DNA，将全酶拴在DNA上，因而使得DNA复制具有持续性。

α-complementation：α 互补。β-半乳糖苷酶的 ω 肽与其 α 肽在活体细胞内互补产生一种活性酶。构建一个克隆载体，它具有利用自身编码的 α 肽与宿主细胞编码的 ω 肽实现 α 互补的特点。因此，该载体单独就具有 α 互补活性，但在该载体中插入克隆基因片段时就不具有 α 互补活性。

γ complex：γ 复合体。由聚合酶 III 全酶的 γ、δ、δ′、χ 和 ψ 亚基组成的复合体，具有钳子装载活性。

α-CTD：细菌 RNA 聚合酶 α-亚基的羧基末端域。

β-galactosidase：β-半乳糖苷酶。能分解乳糖中两种糖组分间糖苷键的酶。

β-galactoside：β-半乳糖苷。一种复合糖，由半乳糖上的碳 1 原子通过 β 键与另一种复合物（通常为糖）连接而成。

β-galactosidic bond：β-半乳糖苷键。将半乳糖的 1 号碳原子与 β-半乳糖苷中的其他复合物相连接的键。

λ gt11：一种能携带外源 DNA 的 LacZ 基因进入 λ 噬菌体中的插入克隆载体。

β-lactamase：β-内酰胺酶。一类分解氨苄青霉素及相关抗生素并使细菌对抗生素具有抗性的酶。

λ phage（lambda phage）：λ 噬菌体。大肠杆菌的一种温和噬菌体。可进行裂解性或溶源性复制。

λ repressor：λ 阻遏物。形成二聚体结合到 λ 噬菌体操纵子 O_R 和 O_L 上的蛋白质，阻遏除阻遏物基因自身外的其他所有的噬菌体基因的表达。

σ^{43}：枯草芽孢杆菌中主要的 σ 因子。

σ^{70}：大肠杆菌中主要的 σ 因子。

12/23 rule：12/23 规则。一种用于免疫球蛋白和 T 细胞受体基因成熟的重组规则。在此规则中 12 信号总是与 23 信号结合，而相似的信号从不相互结合。

12 signal：12 信号。在免疫球蛋白基因形成机制中，将保守的七聚体和九聚体间隔开的一段非保守的 12bp 的重组信号序列（RSS）。

19S particle：19S 颗粒。蛋白酶体上的调节部分，也能促进某些基因的转录延伸。

23 signal：23 信号。在免疫球蛋白基因形成机制中，将保守的七聚体和九聚体间隔开的一段非保守的 23bp 序列的重组信号序列（RSS）。

2μm plasmid：2μm 质粒。用于酵母克隆的基本质粒。

30S initiation complex：30S 起始复合体。由 30S 核糖体亚基、mRNA、fMet-$tRNA_f^{Met}$、起始因子以及 GTP 组成的复合物。此复合物与 50S 核糖体亚基结合。

30S ribosomal subunit：30S 核糖体亚基。细菌核糖体小亚基，参与 mRNA 的解码。

3′-end：3′端。带有游离的 3′羟基的多聚核苷酸链的末端。

4-thioU（sU）：4-硫尿嘧啶能被（以 4-thioUMP 形式）掺入 RNA 然后通过紫外交联到任何结合于 4-thioUMP 位点的 RNA 结合蛋白上的一种光敏核苷。

50S ribosomal subunit：50S 核糖体亚基。细菌的核糖体大亚基，与肽键的合成有关。

5′-end：5′端。具有一个游离（或磷酸化的或帽子化的）5′-羟基的多聚核苷酸的末端。

[6-4] photoproduct：[6-4] 光生产物。由紫外线照射所产生的 DNA 损伤，其中一个嘧啶的 6 位碳原子与相邻嘧啶的 4 位碳原子间形成共价连接。

70S initiation complex：70S 起始复合体。由 70S 核糖体、mRNA 和 fMet-$tRNA_f^{Met}$ 组成的翻译起始复合体。

7-methyl guanosine（m^7G）：7-甲基鸟嘌呤。在真核生物 mRNA 起始端的帽子核苷酸。

7SL RNA：识别分泌蛋白信号肽的小 RNA。

8-oxoguanine（8-hydroxyguanine；oxoG）：8-羟基鸟嘌呤（oxoG）。在 DNA 分子的鸟嘌呤 8 位上含有一个羟基的氧化产物。

翻译　兰海燕　校译　马　纪　高友军　郑用琏

参 考 文 献

第一章

Creighton, H.B., and B. McClintock. 1931. A correlation of cytological and genetical crossing-over in *Zea mays*. *Proceedings of the National Academy of Sciences* 17:492–97.

Mirsky, A.E. 1968. The discovery of DNA. *Scientific American* 218 (June):78–88.

Morgan, T.H. 1910. Sex-limited inheritance in *Drosophila*. *Science* 32:120–22.

Sturtevant, A.H. 1913. The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. *Journal of Experimental Zoology* 14:43–59.

第二章

Adams, R.L.P., R.H. Burdon, A.M. Campbell, and R.M.S. Smellie, eds. 1976. *Davidson's The Biochemistry of the Nucleic Acids*, 8th ed. The structure of DNA, chapter 5. New York: Academic Press.

Avery, O.T., C.M. McLeod, and M. McCarty. 1944. Studies on the chemical nature of the substance-inducing transformation of pneumococcal types. *Journal of Experimental Medicine* 79:137–58.

Chargaff, E. 1950. Chemical specificity of the nucleic acids and their enzymatic degradation. *Experientia* 6:201–9.

Dickerson, R.E. 1983. The DNA helix and how it reads. *Scientific American* 249 (December): 94–111.

Hershey, A.D., and M. Chase. 1952. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *Journal of General Physiology* 36:39–56.

Watson, J.D., and F.H.C. Crick. 1953. Molecular structure of the nucleic acids: A structure for deoxyribose nucleic acid. *Nature* 171:737–38.

———. 1953. Genetical implications of the structure of deoxyribonucleic acid. *Nature* 171:964–67.

第三章

Beadle, G.W., and E.L. Tatum. 1941. Genetic control of biochemical reactions in *Neurospora*. *Proceedings of the National Academy of Sciences* 27:499–506.

Brenner, S., F. Jacob, and M. Meselson. 1961. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* 190:576–81.

Corwin, H.O., and J.B. Jenkins, eds. 1976. *Conceptual Foundations of Genetics: Selected Readings*. Boston: Houghton Mifflin.

Crick, F.H.C. 1958. On protein synthesis. *Symposium of the Society for Experimental Biology* 12:138–63.

Meselson, M., and F.W. Stahl. 1958. The replication of DNA in *Escherichia coli*. *Proceedings of the National Academy of Sciences* 44:671–82.

第四章

Capecchi, N.R. 1994. Targeted gene replacement. *Scientific American* 270 (March):52–59.

Chilton, M.-D. 1983. A vector for introducing new genes into plants. *Scientific American* 248 (June):50–59.

Cohen, S. 1975. The manipulation of genes. *Scientific American* 233 (July):24–33.

Cohen, S., A. Chang, H. Boyer, and R. Helling. 1973. Construction of biologically functional bacterial plasmids in vitro. *Proceedings of the National Academy of Sciences* 70:3240–44.

Gasser, C.S., and R.T. Fraley. 1992. Transgenic crops. *Scientific American* 266 (June):62–69.

Gilbert, W., and L. Villa-Komaroff. 1980. Useful proteins from recombinant bacteria. *Scientific American* 242 (April):74–94.

Nathans, D., and H.O. Smith. 1975. Restriction endonucleases in the analysis and restructuring of DNA molecules. *Annual Review of Biochemistry* 44:273–93.

Sambrook, J., and D. Russell. 2001. *Molecular Cloning: A Laboratory Manual*, 3rd ed. Plainview, NY: Cold Spring Harbor Laboratory Press.

Watson, J.D., J. Tooze, and D.T. Kurtz. 1983. *Recombinant DNA: A Short Course*. New York: W.H. Freeman.

第五章

Galas, D.J. and A. Schmitz. 1978. DNase footprinting: A simple method for the detection of protein–DNA binding specificity. *Nucleic Acids Research* 5:3157–70.

Lichter, P. 1990. High resolution mapping of human chromosome 11 by in situ hybridization with cosmid clones. *Science* 247:64–69.

Sambrook, J., et al. 2001. *Molecular Cloning: A Laboratory Manual*, 3rd ed. Plainview, NY: Cold Spring Harbor Laboratory Press.

第六章

General References and Reviews

Busby, S. and R.H. Ebright. 1994. Promoter structure, promoter recognition, and transcription activation in prokaryotes. *Cell* 79:743–46.

Geiduschek, E.P. 1997. Paths to activation of transcription. *Science* 275:1614–16.

Helmann, J.D. and M.J. Chamberlin. 1988. Structure and function of bacterial sigma factors. *Annual Review of Biochemistry* 57:839–72.

Landick, R. 1999. Shifting RNA polymerase into overdrive. *Science* 284:598–99.

Landick, R. and J.W. Roberts. 1996. The shrewd grasp of RNA polymerase. *Science* 273:202–3.

Losick, R. and M.J. Chamberlin. 1976. *RNA Polymerase*. Plainview, NY: Cold Spring Harbor Laboratory Press.

Richardson, J.P. 1996. Structural organization of transcription termination factor rho. *Journal of Biological Chemistry* 271:1251–54.

Young, B.A., T.M. Gruber, and C.A. Gross. 2002. Views of transcription initiation. *Cell* 109:417–20.

Research Articles

Bar-Nahum, G. and E. Nudler. 2001. Isolation and characterization of σ^{70}-retaining transcription elongation complexes from *E. coli*. *Cell* 106:443–51.

Bautz, E.K.F., F.A. Bautz, and J.J. Dunn. 1969. *E. coli* σ factor: A positive control element in phage T4 development. *Nature* 223:1022–24.

Blatter, E.E., W. Ross, H. Tang, R.L. Gourse, and R.H. Ebright. 1994. Domain organization of RNA polymerase α subunit: C-terminal 85 amino acids constitute a domain capable of dimerization and DNA binding. *Cell* 78:889–96.

Brennan, C.A., A.J. Dombroski, and T. Platt. 1987. Transcription termination factor rho is an RNA–DNA helicase. *Cell* 48:945–52.

Burgess, R.R., A.A. Travers, J.J. Dunn, and E.K.F. Bautz. 1969. Factor stimulating transcription by RNA polymerase. *Nature* 221:43–46.

Campbell, E.A., N. Korzheva, A. Mustaev, K. Murakami, S. Nair, A. Goldfarb, and S.A. Darst. 2001. Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell* 104:901–12.

Carpousis, A.J. and J.D. Gralla. 1980. Cycling of ribonucleic

acid polymerase to produce oligonucleotides during initiation in vitro at the *lac* UV5 promoter. *Biochemistry* 19:3245–53.
Dombroski, A.J., W.A. Walter, M.T. Record, Jr., D.A. Siegele, and C.A. Gross. (1992). Polypeptides containing highly conserved regions of transcription initiation factor σ^{70} exhibit specificity of binding to promoter DNA. *Cell* 70:501–12.
Farnham, P.J. and T. Platt. 1980. A model for transcription termination suggested by studies on the *trp* attenuator in vitro using base analogs. *Cell* 20:739–48.
Fukuda, R. and A. Ishihama. 1974. Subunits of RNA polymerase in function and structure: V. Maturation in vitro of core enzyme from *Escherichia coli*. *Journal of Molecular Biology* 87:523–40.
Gamper, H. and J. Hearst. 1982. A topological model for transcription based on unwinding angle analysis of *E. coli* RNA polymerase binary, initiation, and ternary complexes. *Cell* 29:81–90.
Grachev, M.A., T.I. Kolocheva, E.A. Lukhtanov, and A.A. Mustaev. 1987. Studies on the functional topography of *Escherichia coli* RNA polymerase: Highly selective affinity labelling of initiating substrates. *European Journal of Biochemistry* 163:113–21.
Hayward, R.S., K. Igarashi, and A. Ishihama. 1991. Functional specialization within the α-subunit of *Escherichia coli* RNA polymerase. *Journal of Molecular Biology* 221:23–29.
Heil, A. and W. Zillig. 1970. Reconstitution of bacterial DNA-dependent RNA polymerase from isolated subunits as a tool for the elucidation of the role of the subunits in transcription. *FEBS Letters* 11:165–71.
Hinkle, D.C. and M.J. Chamberlin. 1972. Studies on the binding of *Escherichia coli* RNA polymerase to DNA: I. The role of sigma subunit in site selection. *Journal of Molecular Biology* 70:157–85.
Hsieh, T. -s. and J.C. Wang. 1978. Physicochemical studies on interactions between DNA and RNA polymerase: Ultraviolet absorbance measurements. *Nucleic Acids Research* 5:3337–45.
Malhotra, A., E. Severinova, and S.A. Darst. 1996. Crystal structure of a σ^{70} subunit fragment from *E. coli* RNA polymerase. *Cell* 87:127–36.
Mukhopadhyay, J., A.N. Kapanidis, V. Mekler, E. Kortkhonjia, Y.W. Ebright, and R.H. Ebright. 2001. Translocation of σ^{70} with RNA polymerase during transcription: Fluorescence resonance energy transfer assay for movement relative to DNA. *Cell* 106:453–63.
Murakami, K.S., S. Masuda, E.A. Campbell, O. Muzzin, and S.A. Darst. 2002. Structural basis of transcription initiation: An RNA polymerase holoenzyme-DNA complex. *Science* 296:1285–90.
Murakami, K.S., S. Masuda, and S.A. Darst. 2002. Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science* 296:1280–84.
Nudler, E., E. Avetissova, V. Makovtsov, and A. Goldfarb. 1996. Transcription processivity: Protein–DNA interactions holding together the elongation complex. *Science* 273:211–17.
Nudler, E., A. Mustaev, E. Lukhtanov, and A. Goldfarb. 1997. The RNA–DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell* 89:33–41.
Paul, B.J., M.M. Barker, W. Ross, D.A. Schneider, C. Webb, J.W. Foster, and R.L. Gourse. 2004. DskA. A critical component of the transcription initiation machinery that potentiates the regulation of rRNA promoters by ppGpp and the initiating NTP. *Cell* 118:311–22.
Roberts, J.W. 1969. Termination factor for RNA synthesis. *Nature* 224:1168–74.
Ross, W., K.K. Gosink, J. Salomon, K. Igarashi, C. Zou, A. Ishihama, K. Severinov, and R.L. Gourse. 1993. A third recognition element in bacterial promoters: DNA binding by the α subunit of RNA polymerase. *Science* 262:1407–13.
Saucier, J. -M. and J.C. Wang. 1972. Angular alteration of the DNA helix by *E. coli* RNA polymerase. *Nature New Biology* 239:167–70.
Sidorenkov, I., N. Komissarova, and M. Kashlev. 1998. Crucial role of the RNA:DNA hybrid in the processivity of transcription. *Molecular Cell* 2:55–64.
Siebenlist, U. 1979. RNA polymerase unwinds an 11-base pair segment of a phage T7 promoter. *Nature* 279:651–52.
Toulokhonov, I., I. Artsimovitch, and R. Landick. 2001. Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins. *Science* 292:730–33.
Travers, A.A. and R.R. Burgess. 1969. Cyclic re-use of the RNA polymerase sigma factor. *Nature* 222:537–40.
Yarnell, W.S. and J.W. Roberts. 1999. Mechanism of intrinsic transcription termination and antitermination. *Science* 284:611–15.
Young, B.A., L.C. Anthony, T.M. Gruber, T.M. Arthur, E. Heyduk, C.Z. Lu, M.M. Sharp, T. Heyduk, R.R. Burgess and C.A. Gross. 2001. A coiled-coil from the RNA polymerase β′ subunit allosterically induces selective nontemplate strand binding by σ^{70}. *Cell* 105:935–44.
Vassylyev, D.G., S.-i Sekine, O. Laptenko, J. Lee, M.N. Vassylyeva, S. Borukhov, and S. Yokoyama. 2002. Crystal structure of bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* 417:712–19.
Zhang, G., E.A. Campbell, L. Minakhin, C. Richter, K. Severinov, and S.A. Darst. 1999. Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* 98:811–24.
Zhang, G. and S.A. Darst. 1998. Structure of the *Escherichia coli* RNA polymerase α subunit amino terminal domain. *Science* 281:262–66.

第七章

General References and Reviews

Beckwith, J.R. and D. Zipser, eds. 1970. *The Lactose Operon*. Plainview, NY: Cold Spring Harbor Laboratory Press.
Corwin, H.O. and J.B. Jenkins. *Conceptual Foundations of Genetics: Selected Readings*. 1976. Boston: Houghton Mifflin Co.
Jacob, F. 1966. Genetics of the bacterial cell (Nobel lecture). *Science* 152:1470–78.
Matthews, K.S. 1996. The whole lactose repressor. *Science* 271:1245–46.
Miller, J.H. and W.S. Reznikoff, eds. 1978. *The Operon*. Plainview, NY: Cold Spring Harbor Laboratory Press.
Monod, J. 1966. From enzymatic adaptation to allosteric transitions (Nobel lecture). *Science* 154:475–83.
Ptashne, M. 1989. How gene activators work. *Scientific American* 260 (January):24–31.
Ptashne, M. and W. Gilbert. 1970. Genetic repressors. *Scientific American* 222 (June):36–44.
Rojo, F. 1999. Repression of transcription initiation in bacteria. *Journal of Bacteriology* 181:2987–91.
Schleif, R. 1987. DNA binding proteins. *Science* 241:1182–87.
Vitreschak, A.G., D.A. Rodionov, A.A. Mironov, and M.S. Gelfand. 2004. Riboswitches: The oldest mechanism for the regulation of gene expression? *Trends in Genetics* 20:44–50.
Winkler, W.C. and R.R. Breaker. 2003. Genetic control by metabolite-binding riboswitches. *Chembiochem* 4:1024–32.

Research Articles

Adhya, S. and S. Garges. 1990. Positive control. *Journal of Biological Chemistry* 265:10797–800.
Benoff, B., H. Yang, C.L. Lawson, G. Parkinson, J. Liu, E. Blatter, Y.W. Ebright, H.M. Berman, and R.H. Ebright. 2002. Structural basis of transcription activation: The CAP–αCTD–DNA complex. *Science* 297:1562–66.
Busby, S. and R.H. Ebright. 1994. Promoter structure, promoter recognition, and transcription activation in prokaryotes. *Cell* 79:743–46.

Chen, B., B. deCrombrugge, W.B. Anderson, M.E. Gottesman, I. Pastan, and R.L. Perlman. 1971. On the mechanism of action of *lac* repressor. *Nature New Biology* 233:67–70.
Chen, Y., Y.W. Ebright, and R.H. Ebright. 1994. Identification of the target of a transcription activator protein by protein–protein photocrosslinking. *Science* 265:90–92.
Emmer, M., B. deCrombrugge, I. Pastan, and R. Perlman. 1970. Cyclic-AMP receptor protein of *E. coli*: Its role in the synthesis of inducible enzymes. *Proceedings of the National Academy of Sciences USA* 66:480–87.
Gilbert, W. and B. Müller-Hill. 1966. Isolation of the *lac* repressor. *Proceedings of the National Academy of Sciences USA* 56:1891–98.
Igarashi, K. and A. Ishihama. 1991. Bipartite functional map of the *E. coli* RNA polymerase α subunit: Involvement of the C-terminal region in transcription activation by cAMP–CRP. *Cell* 65:1015–22.
Jacob, F. and J. Monod. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* 3:318–56.
Krummel, B. and M.J. Chamberlin. 1989. RNA chain initiation by *Escherichia coli* RNA polymerase. Structural transitions of the enzyme in the early ternary complexes. *Biochemistry* 28:7829–42.
Lee, J. and A. Goldfarb. 1991. *Lac* repressor acts by modifying the initial transcribing complex so that it cannot leave the promoter. *Cell* 66:793–98.
Lewis, M., G. Chang, N.C. Horton, M.A. Kercher, H.C. Pace, M.A. Schumacher, R.G. Brennan, and P. Lu. 1996. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* 271:1247–54.
Lobell, R.B. and R.F. Schleif. 1991. DNA looping and unlooping by AraC protein. *Science* 250:528–32.
Malan, T.P. and W.R. McClure. 1984. Dual promoter control of the *Escherichia coli* lactose operon. *Cell* 39:173–80.
Oehler, S., E.R. Eismann, H. Krämer, and B. Müller-Hill. 1990. The three operators of the *lac* operon cooperate in repression. *The EMBO Journal* 9:973–79.
Schlax, P.J., M.W. Capp, and M.T. Record, Jr. 1995. Inhibition of transcription initiation by *lac* repressor. *Journal of Molecular Biology* 245:331–50.
Schultz, S.C., G.C. Shields, and T.A. Steitz. 1991. Crystal structure of a CAP–DNA complex: The DNA is bent by 90 degrees. *Science* 253:1001–7.
Straney, S. and D.M. Crothers. 1987. *Lac* repressor is a transient gene-activating protein. *Cell* 51:699–707.
Winkler, W.C., S. Cohen-Chalamish, and R.R. Breaker. 2002. An mRNA structure that controls gene expression by binding FMN. *Proceedings of the National Academy of Sciences, USA* 99:15908–13.
Wu, H.-M. and D.M. Crothers. 1984. The locus of sequence-directed and protein-induced DNA bending. *Nature* 308:509–13.
Yanofsky, C. 1981. Attenuation in the control of expression of bacterial operons. *Nature* 289:751–58.
Zubay, G., D. Schwartz, and J. Beckwith. 1970. Mechanism of activation of catabolite-sensitive genes: A positive control system. *Proceedings of the National Academy of Sciences USA* 66:104–10.

第八章

General References and Reviews

Busby, S. and R.H. Ebright. 1994. Promoter structure, promoter recognition, and transcription activation in prokaryotes. *Cell* 79:743–46.
Goodrich, J.A. and W.R. McClure. 1991. Competing promoters in prokaryotic transcription. *Trends in Biochemical Sciences* 15:394–97.
Gralla, J.D. 1991. Transcriptional control—lessons from an *E. coli* data base. *Cell* 66:415–18.
Greenblatt, J., J.R. Nodwell, and S.W. Mason. 1993. Transcriptional antitermination. *Nature* 364:401–6.
Helmann, J.D. and M.J. Chamberlin. 1988. Structure and function of bacterial sigma factors. *Annual Review of Biochemistry* 57:839–72.
Ptashne, M. 1992. *A Genetic Switch*. Cambridge, MA: Cell Press.

Research Articles

Dodd, I.B., A.J. Perkins, D. Tsemitsidis, and J.B. Egan. 2001. Octamerization of λ CI repressor is needed for effective repression of P_{RM} and efficient switching from lysogeny. *Genes and Development* 15:3013–21.
Gusarov, I. and E. Nudler. 1999. The mechanism of intrinsic transcription termination. *Molecular Cell* 3:495–504.
Gusarov, I. and E. Nudler. 2001. Control of intrinsic transcription termination by N and NusA: The basic mechanisms. *Cell* 107:437–49.
Haldenwang, W.G., N. Lang, and R. Losick. 1981. A sporulation-induced sigma-like regulatory protein from *B. subtilis*. *Cell* 23:615–24.
Hawley, D.K. and W.R. McClure. 1982. Mechanism of activation of transcription initiation from the λ P_{RM} promoter. *Journal of Molecular Biology* 157:493–525.
Ho, Y.-S., D.L. Wulff, and M. Rosenberg. 1983. Bacteriophage λ protein cII binds promoters on the opposite face of the DNA helix from RNA polymerase. *Nature* 304:703–8.
Johnson, W.C., C.P. Moran, Jr., and R. Losick. 1983. Two RNA polymerase sigma factors from *Bacillus subtilis* discriminate between overlapping promoters for a developmentally regulated gene. *Nature* 302:800–4.
Li, M., H. Moyle, and M.M. Suskind. 1994. Target of the transcriptional activation function of phage λ cI protein. *Science* 263:75–77.
Meyer, B.J., D.G. Kleid, and M. Ptashne. 1975. λ repressor turns off transcription of its own gene. *Proceedings of the National Academy of Science, USA* 72:4785–89.
Pero, J., R. Tjian, J. Nelson, and R. Losick. 1975. In vitro transcription of a late class of phage SPO1 genes. *Nature* 257:248–51.
Rong, S., M.S. Rosenkrantz, and A.L. Sonenshein. 1986. Transcriptional control of the *B. subtilis spoIID* gene. *Journal of Bacteriology* 165:771–79.
Stragier, P., B. Kunkel, L. Kroos, and R. Losick. 1989. Chromosomal rearrangement generating a composite gene for a developmental transcription factor. *Science* 243:507–12.
Toulokhonov, I., I. Artsimovitch, and R. Landick. 2001. Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins. *Science* 292:730–33.

第九章

General References and Reviews

Geiduschek, E.P. 1997. Paths to activation of transcription. *Science* 275:1614–16.
Kustu, S., A.K. North, and D.S. Weiss. 1991. Prokaryotic transcriptional enhancers and enhancer-binding proteins. *Trends in Biochemical Sciences* 16:397–402.
Schleif, R. 1988. DNA binding by proteins. *Science* 241:1182–87.

Research Articles

Aggarwal, A.K., D.W. Rodgers, M. Drottar, M. Ptashne, and S.C. Harrison. 1988. Recognition of a DNA operator by the repressor of phage 434: A view at high resolution. *Science* 242:899–907.
Griffith, J., A. Hochschild, and M. Ptashne. 1986. DNA loops induced by cooperative binding of λ repressor. *Nature* 322:750–52.

Herendeen, D.R., G.A. Kassavetis, J. Barry, B.M. Alberts, and E.P. Geiduschek. 1990. Enhancement of bacteriophage T4 late transcription by components of the T4 DNA replication apparatus. *Science* 245:952–58.
Hochschild, A., J. Douhann III, and M. Ptashne. 1986. How λ repressor and λ cro distinguish between O_R1 and O_R3. *Cell* 47:807–16.
Hochschild, A. and M. Ptashne. 1986. Cooperative binding of λ repressors to sites separated by integral turns of the DNA helix. *Cell* 44:681–87.
Jordan, S.R. and C.O. Pabo. 1988. Structure of the lambda complex at 2.5 Å resolution: Details of the repressor–operator interactions. *Science* 242:893–99.
Popham, D.L., D. Szeto, J. Keener, and S. Kustu. 1989. Function of a bacterial activator protein that binds to transcriptional enhancers. *Science* 243:629–35.
Sauer, R.T., R.R. Yocum, R.F. Doolittle, M. Lewis, and C.O. Pabo. 1982. Homology among DNA-binding proteins suggests use of a conserved super-secondary structure. *Nature* 298:447–51.
Schevitz, R.W., Z. Otwinowski, A. Joachimiak, C.L. Lawson, and P. B. Sigler. 1985. The three-dimensional structure of *trp* repressor. *Nature* 317:782–86.
Su, W., S. Porter, S. Kustu, and H. Echols. 1990. DNA looping and enhancer activity: Association between DNA-bound NtrC activator and RNA polymerase at the bacterial *glnA* promoter. *Proceedings of the National Academy of Sciences USA* 87:5504–8.
Wharton, R.P. and M. Ptashne. 1985. Changing the binding specificity of a repressor by redesigning an α-helix. *Nature* 316:601–5.
Zhang, R.-g., A. Joachimiak, C.L. Lawson, R.W. Schevitz, Z. Otwinowski, and P.B. Sigler. 1987. The crystal structure of *trp* aporepressor at 1.8 Å shows how binding tryptophan enhances DNA affinity. *Nature* 327:591–97.

第十章

General References and Reviews

Corden, J.L. 1990. Tales of RNA polymerase II. *Trends in Biochemical Sciences* 15:383–87.
Klug, A. 2001. A marvelous machine for making messages. *Science* 292:1844–46.
Landick, R. 2004. Active-site dynamics in RNA polymerases. *Cell* 116:351–53.
Paule, M.R. and R.J. White. 2000. Transcription by RNA polymerases I and III. *Nucleic Acids Research* 28:1283–98.
Sentenac, A. 1985. Eukaryotic RNA polymerases. *CRC Critical Reviews in Biochemistry* 18:31–90.
Woychik, N.A. and R.A. Young. 1990. RNA polymerase II: Subunit structure and function. *Trends in Biochemical Sciences* 15:347–51.

Research Articles

Benoist, C. and P. Chambon. 1981. In vivo sequence requirements of the SV40 early promoter region. *Nature* 290:304–10.
Bogenhagen, D.F., S. Sakonju, and D.D. Brown. 1980. A control region in the center of the 5S RNA gene directs specific initiation of transcription: II. The 3′ border of the region. *Cell* 19:27–35.
Bushnell, D.A., P. Cramer, and R.D. Kornberg. 2002. Structural basis of transcription: α-amanitin–RNA polymerase cocrystal at 2.8 Å resolution. *Proceedings of the National Academy of Sciences USA* 99:1218–22.
Cramer, P., D.A. Bushnell, and R.D. Kornberg. 2001. Structural basis of transcription: RNA polymerase II at 2.8 Ångstrom resolution. *Science* 292:1863–76.
Das, G., D. Henning, D. Wright, and R. Reddy. 1988. Upstream regulatory elements are necessary and sufficient for transcription of a U6 RNA gene by RNA polymerase III. *EMBO Journal* 7:503–12.
Gillies, S.D., S.L. Morrison, V.T. Oi, and S. Tonegawa. 1983. A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell* 33:717–28.
Gnatt, A.L., P. Cramer, J. Fu, D.A. Bushnell, and R.D. Kornberg. 2001. Structural basis of transcription: An RNA polymerase II elongation complex at 3.3 Å resolution. *Science* 292:1876–82.
Haltiner, M.M., S.T. Smale, and R. Tjian. 1986. Two distinct promoter elements in the human rRNA gene identified by linker scanning mutagenesis. *Molecular and Cellular Biology* 6:227–35.
Kolodziej, P.A., N. Woychik, S.-M. Liao, and R. Young. 1990. RNA polymerase II subunit composition, stoichiometry, and phosphorylation. *Molecular and Cellular Biology* 10:1915–20.
Kutach, A.K. and J.T. Kadonaga. 2000. The downstream promoter element DPE appears to be as widely used as the TATA box in *Drosophila* core promoters. *Molecular and Cellular Biology* 20:4754–64.
Learned, R.M., T.K. Learned, M.M. Haltiner, and R.T. Tjian. 1986. Human rRNA transcription is modulated by the coordinate binding of two factors to an upstream control element. *Cell* 45:847–57.
Lee, T.I. and R.A. Young. 2000. Eukaryotic transcription. *Annual Review of Genetics* 34:77–137.
McKnight, S.L. and R. Kingsbury. 1982. Transcription control signals of a eukaryotic protein-coding gene. *Science* 217:316–24.
Murphy, S., C. Di Liegro, and M. Melli. 1987. The in vitro transcription of the 7SK RNA gene by RNA polymerase III is dependent only on the presence of an upstream promoter. *Cell* 51:81–87.
Pieler, T., J. Hamm, and R.G. Roeder. 1987. The 5S gene internal control region is composed of three distinct sequence elements, organized as two functional domains with variable spacing. *Cell* 48:91–100.
Roeder, R.G. and W.J. Rutter. 1969. Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature* 224:234–37.
Roeder, R.G. and W.J. Rutter. 1970. Specific nucleolar and nucleoplasmic RNA polymerases. *Proceedings of the National Academy of Sciences USA* 65:675–82.
Sakonju, S., D.F. Bogenhagen, and D.D. Brown. 1980. A control region in the center of the 5S RNA gene directs initiation of transcription: I. The 5′ border of the region. *Cell* 19:13–25.
Sayre, M.H., H. Tschochner, and R.D. Kornberg. 1992. Reconstitution of transcription with five purified initiation factors and RNA polymerase II from *Saccharomyces cerevisiae*. *Journal of Biological Chemistry* 267:23376–82.
Sklar, V.E.F., L.B. Schwartz, and R.G. Roeder. 1975. Distinct molecular structures of nuclear class I, II, and III DNA-dependent RNA polymerases. *Proceedings of the National Academy of Sciences USA* 72:348–52.
Smale, S.T. and D. Baltimore. 1989. The "initiator" as a transcription control element. *Cell* 57:103–13.
Ullu, E. and A.M. Weiner. 1985. Upstream sequences modulate the internal promoter of the human 7SL RNA gene. *Nature* 318:371–74.
Weinman, R. and R.G. Roeder. 1974. "Role of DNA-dependent RNA polymerase III in the transcription of the tRNA and 5S rRNA genes." *Proceedings of the National Academy of Sciences USA* 71:1790–94.
Westover, K.D., D.A. Bushnell, and R.D. Kornberg. 2004. Structural basis of transcription: Separation of RNA from DNA by RNA polymerase II. *Science* 303:1014–16.
Westover, K.D., D.A. Bushnell, and R.D. Kornberg. 2004. Structural basis of transcription: Nucleotide selection by rotation in the RNA polymerase II active center. *Cell* 119:481–89.
Woychik, N.A., S.M. Liao, P.A. Kolodziej, and R.A. Young. 1990. Subunits shared by eukaryotic nuclear RNA polymerases. *Genes and Development* 4:313–23.

Woychik, N.A., et al. 1993. Yeast RNA polymerase II subunit RPB11 is related to a subunit shared by RNA polymerase I and III. *Gene Expression* 3:77–82.

第十一章

General References and Reviews

Asturias, F.J. and J.L. Craighead. 2003. RNA polymerase II at initiation. *Proceedings of the National Academy of Sciences USA*. 100:6893–95.

Berk, A.J. 2000. TBP-like factors come into focus. *Cell* 103:5–8.

Buratowski, S. 1997. Multiple TATA-binding factors come back into style. *Cell* 91:13–15.

Burley, S.K. and R.G. Roeder. 1996. Biochemistry and structural biology of transcription factor IID (TFIID). *Annual Review of Biochemistry* 65:769–99.

Chao, D.M. and R.A. Young. 1996. Activation without a vital ingredient. *Nature* 383:119–20.

Conaway, R.C., S.E. Kong, and J.W. Conaway. 2003. TFIIS and GreB: Two like-minded transcription elongation factors with sticky fingers. *Cell* 114:272–74.

Goodrich, J.A., G. Cutler, and R. Tjian. 1996. Contacts in context: Promoter specificity and macromolecular interactions in transcription. *Cell* 84:825–30.

Grant, P. and J.L. Workman. 1998. A lesson in sharing? *Nature* 396:410–11.

Green, M.A. 1992. Transcriptional transgressions. *Nature* 357:364–65.

Hahn, S. 1998. The role of TAFs in RNA polymerase II transcription. *Cell* 95:579–82.

Hahn, S. 2004. Structure and mechanism of the RNA polymerase II transcription machinery. *Nature Structural & Molecular Biology* 11:394–403.

Klug, A. 1993. Opening the gateway. *Nature* 365:486–87.

Paule, M.R. and R.J. White. 2000. Transcription by RNA polymerases I and III. *Nucleic Acids Research* 28:1283–98.

Sharp, P.A. 1992. TATA-binding protein is a classless factor. *Cell* 68:819–21.

Tansey, W.P. and W. Herr. 1997. TAFs: Guilt by association? *Cell* 88:729–32.

White, R.J. and S.P. Jackson. 1992. The TATA-binding protein: A central role in transcription by RNA polymerases I, II, and III. *Trends in Genetics* 8:284–88.

Research Articles

Armache, K.-J., H. Kettenberger, and P. Cramer. 2003. Architecture of initiation-competent 12-subunit RNA polymerase II. *Proceedings of the National Academy of Sciences USA*. 100:6964–68.

Bell, S.P., R.M. Learned, H.-M. Jantzen, and R. Tjian. 1988. Functional cooperativity between transcription factors UBF1 and SL1 mediates human ribosomal RNA synthesis. *Science* 241:1192–97.

Brand, M., C. Leurent, V. Mallouh, L. Tora, and P. Schultz. 1999. Three-dimensional structures of the TAF_{II}-containing complexes TFIID and TFTC. *Science* 286:2151–53.

Bushnell, D.A. and R.D. Kornberg. 2003. Complete, 12-subunit RNA polymerase II at 4.1-Å resolution: Implications for the initiation of transcription. *Proceedings of the National Academy of Sciences USA*. 100:6969–73.

Bushnell, D.A., K.D. Westover, R.E. Davis, and R.D. Kornberg. 2004. Stuctural basis of transcription: An RNA polymerase II-TFIIB cocrystal at 4.5 angstroms. *Science* 303:983–88.

Chen, H.-T. and S. Hahn. 2004. Mapping the location of TFIIB within the RNA polymerase II transcription preinitiation complex: A model for the structure of the PIC. *Cell* 119:169–80.

Chen, J.L., L.D. Attardi, C.P. Verrijzer, K. Yokomori, and R. Tjian. 1994. Assembly of recombinant TFIID reveals differential coactivator requirements for distinct transcriptional activators. *Cell* 79:93–105.

Dynlacht, B.D., T. Hoey, and R. Tjian. (1991). Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. *Cell* 66:563–76.

Flores, O., H. Lu, M. Killeen, J. Greenblatt, Z.F. Burton, and D. Reinberg. 1991. The small subunit of transcription factor IIF recruits RNA polymerase II into the preinitiation complex. *Proceedings of the National Academy of Sciences USA* 88:9999–10003.

Flores, O., E. Maldonado, and D. Reinberg. 1989. Factors involved in specific transcription by mammalian RNA polymerase II: Factors IIE and IIF independently interact with RNA polymerase II. *Journal of Biological Chemistry* 264:8913–21.

Guzder, S.N., P. Sung, V. Bailly, L. Prakash, and S. Prakash. 1994. RAD25 is a DNA helicase required for DNA repair and RNA polymerase II transcription. *Nature* 369:578–81.

Hansen, S.K., S. Takada, R.H. Jacobson, J.T. Lis, and R. Tjian. 1997. Transcription properties of a cell type-specific TATA binding protein, TRF. *Cell* 91:71–83.

Holmes, M.C. and R. Tjian. 2000. Promoter-selective properties of the TBP-related factor TRF1. *Science* 288:867–70.

Holstege, F.C.P., E.G. Jennings, J.J. Wyrick, T.I. Lee, C.J. Hengartner, M.R. Green, T.R. Golub, E.S. Lander, and R.A. Young. 1998. Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* 95:717–28.

Honda, B.M. and R.G. Roeder. 1980. Association of a 5S gene transcription factor with 5S RNA and altered levels of the factor during cell differentiation. *Cell* 22:119–26.

Kassavetis, G.A., B.R. Braun, L.H. Nguyen, and E.P. Geiduschek. 1990. *S. cerevisiae* TFIIIB is the transcription initiation factor proper of RNA polymerase III, while TFIIIA and TFIIIC are assembly factors. *Cell* 60:235–45.

Kassavetis, G.A., D.L. Riggs, R. Negri, L.H. Nguyen, and E.P. Geiduschek. 1989. Transcription factor IIIB generates extended DNA interactions in RNA polymerase III transcription complexes on tRNA genes. *Molecular and Cellular Biology* 9:2551–66.

Kettenberger, H., K.-J. Armache, and P. Cramer. 2003. Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell* 114:347–57.

Kim, J.L., D.B. Nikolov, and S.K. Burley. 1993. Co-crystal structure of a TBP recognizing the minor groove of a TATA element. *Nature* 365:520–27.

Kim, T.-K., R.H. Ebright, and D. Reinberg. 2000. Mechanism of ATP-dependent promoter melting by transcription factor IIH. *Science* 288:1418–21.

Kim, Y.J., S. Björklund, Y. Li, M.H. Sayre, and R.D. Kornberg. 1994. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* 77:599–608.

Koleske, A.J. and R.A. Young. 1994. An RNA polymerase II holoenzyme responsive to activators. *Nature* 368:466–69.

Kownin, P., E. Bateman, and M.R. Paule. 1987. Eukaryotic RNA polymerase I promoter binding is directed by protein contacts with transcription initiation factor and is DNA sequence-independent. *Cell* 50:693–99.

Learned, R.M., S. Cordes, and R. Tjian. 1985. Purification and characterization of a transcription factor that confers promoter specificity to human RNA polymerase I. *Molecular and Cellular Biology* 5:1358–69.

Lobo, S.L., M. Tanaka, M.L. Sullivan, and N. Hernandez. 1992. A TBP complex essential for transcription from TATA-less but not TATA-containing RNA polymerase III promoters is part of the TFIIIB fraction. *Cell* 71:1029–40.

Lu, H., L. Zawel, L. Fisher, J.-M. Egly, and D. Reinberg. 1992. Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II. *Nature* 358:641–45.

Maldonado, E., I. Ha, P. Cortes, L. Weis, and D. Reinberg. 1990. Factors involved in specific transcription by mammalian RNA polymerase II: Role of transcription factors IIA, IID, and IIB during formation of a transcription-competent complex. *Molecular and Cellular Biology*

10:6335–47.
Ossipow, V., J.-P. Tassan, E.I. Nigg, and U. Schibler. 1995. A mammalian RNA polymerase II holoenzyme containing all components required for promoter-specific transcription initiation. *Cell* 83:137–46.
Pelham, H.B., Wormington, W.M., and D.D. Brown. 1981. Related 5S rRNA transcription factors in *Xenopus* oocytes and somatic cells. *Proceeding of the National Academy of Sciences USA* 78:1760–64.
Pugh, B.F. and R. Tjian. 1991. Transcription from a TATA-less promoter requires a multisubunit TFIID complex. *Genes and Development* 5:1935–45.
Rowlands, T., P. Baumann, and S.P. Jackson. 1994. The TATA-binding protein: A general transcription factor in eukaryotes and archaebacteria. *Science* 264:1326–29.
Sauer, F., D.A. Wassarman, G.M. Rubin, and R. Tjian. 1996. TAF_{II}s mediate activation of transcription in the *Drosophila* embryo. *Cell* 87:1271–84.
Schultz, M.C., R.H. Roeder, and S. Hahn. 1992. Variants of the TATA-binding protein can distinguish subsets of RNA polymerase I, II, and III promoters. *Cell* 69:697–702.
Setzer, D.R. and D.D. Brown. 1985. Formation and stability of the 5S RNA transcription complex. *Journal of Biological Chemistry* 260:2483–92.
Shastry, B.S., S.-Y. Ng, and R.G. Roeder. 1982. Multiple factors involved in the transcription of class III genes in *Xenopus laevis*. *Journal of Biological Chemistry* 257:12979–86.
Starr, D.B. and D.K. Hawley. 1991. TFIID binds in the minor groove of the TATA box. *Cell* 67:1231–40.
Taggart, K.P., J.S. Fisher, and B.F. Pugh. 1992. The TATA-binding protein and associated factors are components of Pol III transcription factor TFIIIB. *Cell* 71:1051–28.
Takada, S., J.T. Lis, S. Zhou, and R. Tjian. 2000. A TRF1:BRF complex directs *Drosophila* RNA polymerase III transcription. *Cell* 101:459–69.
Tanese, N. 1991. Coactivators for a proline-rich activator purified from the multisubunit human TFIID complex. *Genes and Development* 5:2212–24.
Thomas, M.J., A.A. Platas, and D.K. Hawley. 1998. Transcriptional fidelity and proofreading by RNA polymerase II. *Cell* 93:627–37.
Verrijzer, C.P., J.-L. Chen, K. Yokomori, and R. Tjian. 1995. Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. *Cell* 81:1115–25.
Walker, S.S., J.C. Reese, L.M. Apone, and M.R. Green. 1996. Transcription activation in cells lacking TAF_{II}s. *Nature* 383:185–88.
Wieczorek, E., M. Brand, X. Jacq, and L. Tora. 1998. Function of TAF_{II}-containing complex without TBP in transcription by RNA polymerase II. *Nature* 393:187–91.

第十二章

General References and Reviews

Bell, A.C. and G. Felsenfeld. 1999. Stopped at the border: boundaries and insulators. *Current Opinion in Genetics and Development* 9:191–98.
Blackwood, E.M. and J.T. Kadonaga. 1998. Going the distance: A current view of enhancer action. *Science* 281:60–63.
Carey, M. 1994. Simplifying the complex. *Nature* 368:402–3.
Conaway, R.C., C.S. Brower, and J.W. Conaway. 2002. Emerging roles of ubiquitin in transcription regulation. *Science* 296:1254–58.
Freiman, R.N. and R. Tjian. 2003. Regulating the regulators: Lysine modifications make their mark. *Cell* 112:11–17.
Goodrich, J.A., G. Cutler, and R. Tjian. 1996. Contacts in context: Promoter specificity and macromolecular interactions in transcription. *Cell* 84:825–30.
Hampsey, M. and D. Reinberg. 1999. RNA polymerase II as a control panel for multiple coactivator complexes. *Current Opinion in Genetics and Development* 9:132–39.
Janknecht, R. and T. Hunter. 1996. Transcription. A growing coactivator network. *Nature* 383:22–23.
Montminy, M. 1997. Something new to hang your hat on. *Nature* 387:654–55.
Myer, V.E. and R.A. Young. 1998. RNA polymerase II holoenzymes and subcomplexes. *Journal of Biological Chemistry* 273:27757–60.
Nordheim, A. 1994. CREB takes CBP to tango. *Nature* 370:177–78.
Ptashne, M. and A. Gann. 1997. Transcriptional activation by recruitment. *Nature* 386:569–77.
Roush, W. 1996. "Smart" genes use many cues to set cell fate. *Science* 272:652–53.
Sauer, F. and R. Tjian. 1997. Mechanisms of transcription activation: Differences and similarities between yeast, *Drosophila*, and man. *Current Opinion in Genetics and Development* 7:176–81.
Werner, M.H. and S.K. Burley. 1997. Architectural transcription factors: Proteins that remodel DNA. *Cell* 88:733–36.
West, A.G., M. Gaszner, and G. Felsenfeld. 2002. Insulators: many functions, many mechanisms. *Genes and Development* 16:271–88.

Research Articles

Barberis, A., J. Pearlberg, N. Simkovich, S. Farrell, P. Reinagle, C. Bamdad, G. Sigal, and M. Ptashne. 1995. Contact with a component of the polymerase II holoenzyme suffices for gene activation. *Cell* 81:359–68.
Borggrefe, T., R. Davis, A. Bareket-Samish, and R.D. Kornberg. 2001. Quantitation of the RNA polymerase II transcription machinery in yeast. *Journal of Biological Chemistry* 276:47150–53.
Brent, R. and M. Ptashne. 1985. A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. *Cell* 43:729–36.
Cai, H.N. and P. Shen 2001. Effects of *cis* arrangement of chromatin insulators on enhancer-blocking activity. *Science* 291:493–95.
Dunaway, M. and P. Dröge. 1989. Transactivation of the *Xenopus* rRNA gene promoter by its enhancer. *Nature* 341:657–59.
Ellenberger, T.E., C.J. Brandl, K. Struhl, and S.C. Harrison. 1992. The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted α helices: Crystal structure of the protein–DNA complex. *Cell* 71:1223–37.
Flanagan, P.M., R.J. Kelleher, 3rd, M.H. Sayre, H. Tschochner, and R.D. Kornberg. 1991. A mediator required for activation of RNA polymerase II transcription in vitro. *Nature* 350:436–38.
Geise, K., J. Cox, and R. Grosschedl. 1992. The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. *Cell* 69:185–95.
Kissinger, C.R., B. Liu, E. Martin-Blanco, T.B. Kornberg, and C.O. Pabo. 1990. Crystal structure of an engrailed homeodomain–DNA complex at 2.8 Å resolution: A framework for understanding homeodomain–DNA interactions. *Cell* 63:579–90.
Koleske, A.J. and R.A. Young. 1994. An RNA polymerase II holoenzyme responsive to activators. *Nature* 368:466–69.
Kouzarides, T. and E. Ziff. 1988. The role of the leucine zipper in the fos–jun interaction. *Nature* 336:646–51.
Krebs, J.E. and Dunaway, M. 1998. The scs and scs′ elements impart a *cis* requirement on enhancer–promoter interactions. *Molecular Cell* 1:301–08.
Lee, M.S. 1989. Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science* 245:635–37.
Leuther, K.K., J.M. Salmeron, and S.A. Johnston. 1993. Genetic evidence that an activation domain of GAL4 does not require acidity and may form a β sheet. *Cell* 72:575–85.
Luisi, B.F., W.X. Xu, Z. Otwinowski, L.P. Freedman, K.R. Yamamoto, and P.B. Sigler. 1991. Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* 352:497–505.

Ma, P.C.M., M.A. Rould, H. Weintraub, and C.O. Pabo. 1994. Crystal structure of MyoD bHLH domain–DNA complex: Perspectives on DNA recognition and implications for transcription activation. *Cell* 77:451–59.
Marmorstein, R., M. Carey, M. Ptashne, and S.C. Harrison. 1992. DNA recognition by GAL4: Structure of a protein–DNA complex. *Nature* 356:408–14.
Muravyova, E., A. Golovnin, E. Gracheva, A. Parshikov, T. Belenkaya, V. Pirrotta, and P. Georgiev. 2001. Loss of insulator activity by paired Su(Hw) chromatin insulators. *Science* 291:495–98.
O'Shea, E.K., J.D. Klemm, P.S. Kim, and T. Alber. 1991. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science* 254:539–44.
Pavletich, N.P. and C.O. Pabo. 1991. Zinc finger–DNA recognition: Crystal structure of a Zif 268–DNA complex at 2.1 Å. *Science* 252:809–17.
Romano, L.A. and G.A. Wray. 2003. Conservation of *Endo 16* expression in sea urchins despite evolutionary divergence in both cis and trans-acting components of transcriptional regulation. *Development* 130:4187–99.
Ryu, L., L. Zhou, A.G. Ladurner, and R. Tjian. 1999. The transcriptional cofactor complex CRSP is required for activity of the enhancer-binding protein Sp1. *Nature* 397:446–50.
Stringer, K.F., C. J. Ingles, and J. Greenblatt. 1990. Direct and selective binding of an acidic transcriptional activation domain to the TATA-box factor TFIID. *Nature* 345:783–86.
Thanos, D. and T. Maniatis. 1995. Virus induction of human IFN-β gene expression requires the assembly of an enhanceosome. *Cell* 83:1091–1100.
Xu, W., H. Chen, K. Du, H. Asahara, M. Tini, B.M. Emerson, M. Montminy, and R.M Evans. 2001. A transcriptional switch mediated by cofactor methylation. *Science* 294:2507–11.
Yuh, C.-H., B. Hamid, and E.H. Davidson. 1998. Genomic *cis*-regulatory logic: Experimental and computational analysis of a sea urchin gene. *Science* 279:1896–1902.
Zhu, J. and M. Levine. 1999. A novel *cis*-regulatory element, the PTS, mediates an antiinsulator activity in the *Drosophila* embryo. *Cell* 99:567–75.

第十三章

General References and Reviews

Brownell, J.E. and C.D. Allis. 1996. Special HATs for special occasions: Linking histone acetylation to chromatin assembly and gene activation. *Current Opinion in Genetics and Development* 6:176–84.
Felsenfeld, G. 1996. Chromatin unfolds. *Cell* 86:13–19.
Grunstein, M. 1998. Yeast heterochromatin: Regulation of its assembly and inheritance by histones. *Cell* 93:325–28.
Narlikar, G.J., H.-Y. Fan, and R.E. Kingston. 2002. Cooperation between complexes that regulate chromatin structure and transcription. *Cell* 108:475–87.
Pazin, M.J. and J.T. Kadonaga. 1997. What's up and down with histone deacetylation and transcription? *Cell* 89:325–28.
Pennisi, E. 1996. Linker histones, DNA's protein custodians, gain new respect. *Science* 274:503–4.
Pennisi, E. 1997. Opening the way to gene activity. *Science* 275:155–57.
Pennisi, E. 2000. Matching the transcription machinery to the right DNA. *Science* 288:1372–73.
Perlman, T. and B. Vennström. 1995. The sound of silence. *Nature* 377:387–88.
Rhodes, D. 1997. The nucleosome core all wrapped up. *Nature* 389:231–33.
Roth, S.Y. and C.D. Allis. 1996. Histone acetylation and chromatin assembly: A single escort, multiple dances? *Cell* 87:5–8.
Svejstrup, J.Q. 2003. Histones face the FACT. *Science* 301:1053–55.
Turner, B.M. 2002. Cellular memory and the histone code. *Cell* 111:285–91.
Wolffe, A.P. 1996. Histone deacetylase: A regulator of transcription. *Science* 272:371–72.
Wolffe, A.P. 1997. Sinful repression. *Nature* 387:16–17.
Wolffe, A.P. and D. Pruss. 1996. Targeting chromatin disruption: Transcription regulators that acetylate histones. *Cell* 84:817–19.
Woodcock, C.L. 2005. A milestone in the odyssey of higher-order chromatin structure. *Nature Structural & Molecular Biology* 12:639–40.
Xu, L., C.K. Glass, and M.G. Rosenfeld. 1999. Coactivator and corepressor complexes in nuclear receptor function. *Current Opinion in Genetics and Development* 9:140–47.

Research Articles

Agalioti, T., G. Chen, and D. Thanos. 2002. Deciphering the transcriptional histone acetylation code for a human gene. *Cell* 111:381–92.
Arents, G. and E.N. Moudrianakis. 1993. Topography of the histone octamer surface: Repeating structural motifs utilized in the docking of nucleosomal DNA. *Proceedings of the National Academy of Sciences USA* 90:10489–93.
Arents, G., R.W. Burlingame, B.-C. Wang, W.E. Love, and E.N. Moudrianakis. 1991. The nucleosomal core histone octamer at 3.1 Å resolution: A tripartite protein assembly and a left-handed superhelix. *Proceedings of the National Academy of Sciences USA* 88:10148–52.
Bannister, A.J., P. Zegerman, J.F. Partridge, E.A. Miska, J.O. Thomas, R.C. Allshire, and T. Kouzarides. 2001. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromodomain. *Nature* 410:120–24.
Belotserkovskaya, R., S. Oh, V.A. Bondarenko, G. Orphanides, V.M. Studitsky, and D. Reinberg. 2003. FACT facilitates transcription-dependent nucleosome alteration. *Science* 301:1090–93.
Benyajati, C. and A. Worcel. 1976. Isolation, characterization, and structure of the folded interphase genome of *Drosophila melanogaster*. *Cell* 9:393–407.
Briggs, S.D., T. Xiao, Z.-W. Sun, J.A. Caldwell, J. Shabanowitz, D.F. Hunt, C.D. Allis, and B.D. Strahl. 2002. *Trans*-histone regulatory pathway in chromatin. *Nature* 418:498.
Brownell, J.E. and C.D. Allis. 1996. An activity gel assay detects a single, catalytically active histone acetyltransferase subunit in *Tetrahymena* macronuclei. *Proceedings of the National Academy of Sciences USA* 92:6364–68.
Brownell, J.E., J. Zhou, T. Ranalli, R. Kobayashi, D.G. Edmondson, S.Y. Roth, and C.D. Allis. 1996. *Tetrahymena* histone acetyltransferase A: A homolog of yeast GCN5p linking histone acetylation to gene activation. *Cell* 84:843–51.
Cosma, M.P., T. Tanaka, and K. Nasmyth. 1999. Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle- and developmentally regulated promoter. *Cell* 97:299–311.
Griffith, J. 1975. Chromatin structure: Deduced from a minichromosome. *Science* 187:1202–3.
Grosveld, F., G.B. van Assendelft, D.R. Greaves, and G. Kollias. 1987. Position-independent, high-level expression of the human β-globin gene in transgenic mice. *Cell* 51:975–85.
Jacobson, R.H., A.G. Ladurner, D.S. King, and R. Tjian. 2000. Structure and function of a human $TAF_{II}250$ double bromodomain module. *Science* 288:1422–28.
Lachner, M., D. O'Carroll, S. Rea, K. Mechtler, and T. Jenuwein. 2001. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* 410:116–120.
Laherty, C.D., W.-M. Yang, J.-M. Sun, J.R. Davie, E. Seto, and R.N. Eisenman. 1997. Histone deacetylases associated with the mSin3 corepressor mediate Mad transcriptional repression. *Cell* 89:349–56.
Laybourn, P.J. and J.T. Kadonaga. 1991. Role of nucleosomal cores and histone H1 in regulation of transcription by RNA

polymerase II. *Science* 254:238–45.
Luger, K., A.W. Mäder, R.K. Richmond, D.F. Sargent, and T.J. Richmond. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389:251–60.
Marsden, M.P.F. and U.K. Laemmli. 1979. Metaphase chromosome structure: Evidence for a radial loop model. *Cell* 17:849–58.
Narlikar, G.J., M.L. Phelan, and R.E. Kingston. 2001. Generation and interconversion of multiple distinct nucleosomal states as a mechanism for catalyzing chromatin fluidity. *Molecular Cell* 8:1219–30.
Ogryzko, V.V., R.L. Schiltz, V. Russanova, B.H. Howard, and Y. Nakatani. 1996. The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell* 87:953–59.
Panyim, S. and R. Chalkley. 1969. High resolution acrylamide gel electrophoresis of histones. *Archives of Biochemistry and Biophysics* 130:337–46.
Paulson, J.R. and U.K. Laemmli. 1977. The structure of histone-depleted interphase chromosomes. *Cell* 12:817–28.
Saragosti, S., G. Moyne, and M. Yaniv. 1980. Absence of nucleosomes in a fraction of SV40 chromatin between the origin of replication and the region coding for the late leader RNA. *Cell* 20:65–73.
Saunders, A., J. Werner, E.D. Andrulis, T. Nakayama, S. Hirose, D. Reinberg, and J.T. Lis. 2003. Tracking FACT and RNA polymerase II elongation complex through chromatin in vivo. *Science* 301:1094–96.
Schalch, T., S. Duda, D.F. Sargent, and T.J. Richmond. 2005. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* 436:138–41.
Schlissel, M.S. and D.D. Brown. 1984. The transcriptional regulation of *Xenopus* 5S RNA genes in chromatin: The roles of active stable transcription complexes and histone H1. *Cell* 37:903–13.
Taunton, J., C.A. Hassig, and S.L. Schreiber. 1996. A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science* 272:408–11.
Thoma, F., T. Koller, and A. Klug. 1979. Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructure of chromatin. *Journal of Cell Biology* 83:403–27.

第十四章

General References and Reviews

Baker, B.S. 1989. Sex in flies: The splice of life. *Nature* 340:521–24.
Black, D.L. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annual Review of Biochemistry* 72:291–336.
Fu, X.-D. 2004. Towards a splicing code. *Cell* 119:736–38.
Guthrie, C. 1991. Messenger RNA splicing in yeast: Clues to why the spliceosome is a ribonucleoprotein. *Science* 253:157–63.
Hirose, Y. and J.L. Manley. 2000. RNA polymerase II and the integration of nuclear events. *Genes and Development* 14:1415–29.
Lamm, G.M. and A.I. Lamond. 1993. Non-snRNP protein splicing factors. *Biochimica et Biophysica Acta* 1173:247–65.
Manley, J. 1993. Question of commitment. *Nature* 365:14.
Murray, H.L. and K.A. Jarrell. 1999. Flipping the switch to an active spliceosome. *Cell* 96:599–602.
Newman, A. 2002. RNA enzymes for RNA splicing. *Nature* 413:695–96.
Nilsen, T.W. 2000. The case for an RNA enzyme. *Nature* 408:782–83.
Orphanides, G. and D. Reinberg. 2002. A unified theory of gene expression. *Cell* 108:439–51.
Proudfoot, N., A. Furger, and M.J. Dye. 2002. Integrating mRNA processing with transcription. *Cell* 108:501–12.
Sharp, P.A. 1994. Split genes and RNA splicing. (Nobel Lecture.) *Cell* 77:805–15.
Villa, T., J.A. Pleiss, and C. Guthrie. 2002. Spliceosomal snRNAs: Mg^{2+}-dependent chemistry at the catalytic core? *Cell* 109:149–52.
Weiner, A.M. 1993. mRNA splicing and autocatalytic introns: Distant cousins or the products of chemical determinism? *Cell* 72:161–64.
Wise, J.A. 1993. Guides to the heart of the spliceosome. *Science* 262:1978–79.

Research Articles

Abovich, N. and M. Rosbash. 1997. Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* 89:403–12.
Berget, S.M., C. Moore, and P. Sharp. 1977. Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences USA* 74:3171–75.
Berglund, J.A., K. Chua, N. Abovich, R. Reed, and M. Rosbash. 1997. The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* 89:781–87.
Brody, E. and J. Abelson. 1985. The spliceosome: Yeast pre-messenger RNA associated with a 40S complex in a splicing-dependent reaction. *Science* 228:963–67.
Chua, K. and R. Reed. 1999. The RNA splicing factor hSlu7 is required for correct 3′ splice-site choice. *Nature* 402:207–10.
Early, P., J. Rogers, M. Davis, K. Calame, M. Bond, R. Wall, and L. Hood. 1980. Two mRNAs can be produced from a single immunoglobulin γ gene by alternative RNA processing pathways. *Cell* 20:313–19.
Fu, X.-D. 1993. Specific commitment of different pre-mRNAs to splicing by single SR proteins. *Nature* 365:82–85.
Grabowski, P., R.A. Padgett, and P.A. Sharp. 1984. Messenger RNA splicing in vitro: An excised intervening sequence and a potential intermediate. *Cell* 37:415–27.
Grabowski, P., S.R. Seiler, and P.A. Sharp. 1985. A multicomponent complex is involved in the splicing of messenger RNA precursors. *Cell* 42:345–53.
Kruger, K., P.J. Grabowski, A.J. Zaug, J. Sands, D.E. Gottschling, and T.R. Cech. 1982. Self-splicing RNA: Autoexcision and autocylization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* 31:147–57.
Langford, C. and D. Gallwitz. 1983. Evidence for an intron-contained sequence required for the splicing of yeast RNA polymerase II transcripts. *Cell* 33:519–27.
Lesser, C.F. and C. Guthrie. 1993. Mutations in U6 snRNA that alter splice site specificity: Implications for the active site. *Science* 262:1982–88.
Liao, X.C., J. Tang, and M. Rosbash. 1993. An enhancer screen identifies a gene that encodes the yeast U1 snRNP A protein: Implications for snRNP protein function in pre-mRNA splicing. *Genes and Development* 7:419–28.
Parker, R., P.G. Siliciano, and C. Guthrie. 1987. Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell* 49:229–39.
Peebles, C.L., P. Gegenheimer, and J. Abelson. 1983. Precise excision of intervening sequences from precursor tRNAs by a membrane-associated yeast endonuclease. *Cell* 32:525–36.
Ruby, S.W. and J. Abelson. 1988. An early hierarchic role of U1 small nuclear ribonucleoprotein in spliceosome assembly. *Science* 242:1028–35.
Sontheimer, E.J. and J.A. Steitz. 1993. The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science* 262:1989–96.
Staley, J.P. and C. Guthrie. 1999. An RNA switch at the 5′ splice site requires ATP and the DEAD box protein Prp28p. *Molecular Cell* 3:55–64.
Tian, M. and T. Maniatis. 1993. A splicing enhancer complex controls alternative splicing of *doublesex* pre-mRNA. *Cell* 74:105–14.
Tilghman, S.M., P. Curtis, D. Tiemeier, P. Leder, and C. Weissmann. 1978. The intervening sequence of a mouse

β-globin gene is transcribed within the 15S β-globin mRNA precursor. *Proceedings of the National Academy of Sciences USA* 75:1309–13.
Valadkhan, S. and J.L. Manley. 2001. Splicing-related catalysis by protein-free snRNAs. *Nature* 413:701–07.
Wang, Z., M.E. Rolish, G. Yeo, V. Tung, M. Mawson, and C.B. Burge. 2004. Systematic identification and analysis of exonic splicing silencers. *Cell* 119:831–45.
Wassarman, D.A. and J.A. Steitz. 1992. Interactions of small nuclear RNAs with precursor messenger RNA during in vitro splicing. *Science* 257:1918–25.
Yarus, M., I. Illangesekare, and E. Christian. 1991. An axial binding site in the *Tetrahymena* precursor RNA. *Journal of Molecular Biology* 222:995–1012.
Yean, S.-L., G. Wuenschell, J. Termini, and R.-J. Lin. 2001. Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. *Nature* 408:881–84.
Zeng, C. and S.M. Berget. 2000. Participation of the C-terminal domain of RNA polymerase II in exon definition during pre-mRNA splicing. *Molecular and Cellular Biology* 20:8290–8301.
Zhuang, Y. and A.M. Weiner. 1986. A compensatory base change in U1 snRNA suppresses a 5′-splice site mutation. *Cell* 46:827–35.

第十五章

General References and Reviews

Barabino, S.M.L. and W. Keller. 1999. Last but not least: Regulated poly(A) tail formation. *Cell* 99:9–11.
Bentley, D. 1998. A tale of two tails. *Nature* 395:21–22.
Colgan, D.F. and Manley, J.L. 1997. Mechanism and regulation of mRNA polyadenylation. *Genes and Development* 11:2755–66.
Manley, J.L. and Y. Takagaki. 1996. The end of the message—Another link between yeast and mammals. *Science* 274:1481–82.
Orphanides, G. and D. Reinberg. 2002. A unified theory of gene expression. *Cell* 108:439–51.
Proudfoot, N.J. 1996. Ending the message is not so simple. *Cell* 87:779–81.
Proudfoot, N.J. A. Furger, and M.J. Dye. 2002. Integrating mRNA processing with transcription. *Cell* 108:501–12.
Tollervey, D. 2004. Molecular biology: Termination by torpedo. *Nature* 432:456–57.
Wahle, E. and W. Keller. 1996. The biochemistry of polyadenylation. *Trends in Biochemical Sciences* 21:247–51.
Wickens, M. 1990. How the messenger got its tail: Addition of poly(A) in the nucleus. *Trends in Biochemical Sciences* 15:277–81.

Research Articles

Bardwell, V.J., D. Zarkower, M. Edmonds, and M. Wickens. 1990. The enzyme that adds poly(A) to mRNA is a classical poly(A) polymerase. *Molecular and Cellular Biology* 10:846–49.
Barillà, D., B.A. Lee, and N.J. Proudfoot. 2001. Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences USA* 98:445–50.
Birse, C.E., L. Minvielle-Sebastia, B.A. Lee, W. Keller, and N.J. Proudfoot. 1998. Coupling termination of transcription to messenger RNA maturation in yeast. *Science* 280:298–301.
Dye, M.J. and N.J. Proudfoot. 2001. Multiple transcript cleavage precedes polymerase release in termination by RNA polymerase II. *Cell* 105:669–81.
Fitzgerald, M. and T. Shenk. 1981. The sequence 5′-AAUAAA-3′ forms part of the recognition site for polyadenylation of late SV40 mRNAs. *Cell* 24:251–60.
Fox, C.A., M.D. Sheets, and M.P. Wickens. 1989. Poly(A) addition during maturation of frog oocytes: Distinct nuclear and cytoplasmic activities and regulation by the sequence UUUUUAU. *Genes and Development* 3:2151–56.
Furuichi, Y., A. LaFiandra, and A.J. Shatkin. 1977. 5′-terminal structure and mRNA stability. *Nature* 266:235–39.
Furuichi, Y. and K.-I. Miura. 1975. A blocked structure at the 5′-terminus of mRNA from cytoplasmic polyhedrosis virus. *Nature* 253:374–75.
Furuichi, Y., S. Muthukrishnan, J. Tomasz, and A.J. Shatkin. 1976. Mechanism of formation of reovirus mRNA 5′-terminal blocked and methylated sequence m^7GpppG^mpC. *Journal of Biological Chemistry* 251:5043–53.
Gallie, D.R. 1991. The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. *Genes and Development* 5:2108–16.
Gil, A. and N.J. Proudfoot. 1987. Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit β-globin mRNA 3′-end formation. *Cell* 49:399–406.
Hamm, J. and I.W. Mattaj. 1990. Monomethylated cap structures facilitate RNA export from the nucleus. *Cell* 63:109–18.
Hirose, Y. and J.L. Manley. 1998. RNA polymerase II is an essential mRNA polyadenylation factor. *Nature* 395:93–96.
Hofer, E. and J.E. Darnell. 1981. The primary transcription unit of the mouse β-major globin gene. *Cell* 23:585–93.
Huez, G., G. Marbaix, E. Hubert, M. Leclereq, U. Nudel, H. Soreq, R. Salomon, B. Lebleu, M. Revel, and U.Z. Littauer. 1974. Role of the polyadenylate segment in the translation of globin messenger RNA in *Xenopus* oocytes. *Proceedings of the National Academy of Sciences USA* 71:3143–46.
Izaurralde, E., J. Lewis, C. McGuigan, M. Jankowska, E. Darzynkiewicz, and I.W. Mattaj. 1994. A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell* 78:657–68.
Keller, W., S. Bienroth, K.M. Lang, and G. Christofori. 1991. Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3′ processing signal AAUAAA. *EMBO Journal* 10:4241–49.
Kim, M., N.J. Krogan, L. Vasiljeva, O.J. Rando, E. Nedea, J.F. Greenblatt, and S. Buratowski. 2004. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* 432:517–22.
Komarnitsky, P., E.-J. Cho, and S. Buratowski. 2000. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes and Development* 14:2452–60.
McCracken, S., N. Fong, E. Rosonina, K. Yankulov, G. Brothers, D. Siderovski, A. Hessel, S. Foster, Amgen EST Program, S. Shuman, and D.L. Bentley. 1997. 5′-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes and Development* 11:3306–18.
McDonald, C.C., J. Wilusz, and T. Shenk. 1994. The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. *Molecular and Cellular Biology* 14:6647–54.
Morris, D.P. and A.L. Greenleaf. 2000. The splicing factor, Prp40, binds the phosphorylated carboxy-terminal domain of RNA polymerase II. *Journal of Biological Chemistry* 275:39935–43.
Munroe, D. and A. Jacobson. 1990. mRNA poly(A) tail, a 3′ enhancer of a translational initiation. *Molecular and Cellular Biology* 10:3441–55.
Niwa, M. and S. Berget. 1991. Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns. *Genes and Development* 5:2086–95.
Ohno, M., H. Sakamoto, and Y. Shimura. 1987. Preferential excision of the 5′-proximal intron from mRNA precursors with two introns as mediated by the cap structure. *Proceedings of the National Academy of Sciences USA* 84:5187–91.

Raabe, T., F.J. Bollum, and J.L. Manley. 1991. Primary structure and expression of bovine poly(A) polymerase. *Nature* 353:229–34.
Sheets, M.D. and M. Wickens. 1989. Two phases in the addition of a poly(A) tail. *Genes and Development* 3:1401–12.
Sheiness, D. and J.E. Darnell. 1973. Polyadenylic acid segment in mRNA becomes shorter with age. *Nature New Biology* 241:265–68.
Teixeira, A., A. Tahirir-Alaoui, S. West, B. Thomas, A. Ramadass, I. Martianov, M. Dye, W. James, N.J. Proudfoot, and A. Akoulitchev. 2004. Autocatalytic RNA cleavage in the human β-globin pre-mRNA promotes transcription termination. *Nature* 432:526–30.
Wahle, E. 1991. A novel poly(A)-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation. *Cell* 66:759–68.
Wei, C.M. and Moss, B. 1975. Methylated nucleotides block 5′-terminus of vaccinia virus mRNA. *Proceedings of the National Academy of Sciences USA* 72:318–22.
West, S., N. Gromak, and N.J. Proudfoot. 2004. Human 5′→3′ exonuclease Xrn2 promotes transcription termination at cotranscriptional cleavage sites. *Nature* 432:522–25.

第十六章

General References and Reviews

Bass, B.L. 2000. Double-stranded RNA as a template for gene silencing. *Cell* 101:235–38.
Carrington, J.C. and V. Ambros. 2003. Role of microRNAs in plant and animal development. *Science* 301:336–38.
Dernburg, A.F. 2002. A Chromosome RNAissance. *Cell* 11:159–62.
Filipowicz, W. 2005. RNAi: The nuts and bolts of the RISC machine. *Cell* 122:17–20.
Keegan, L.P., A. Gallo, and M.A. O'Connell. 2000. Survival is impossible without an editor. *Science* 290:1707–09.
Nilsen, T.W. 1994. Unusual strategies of gene expression and control in parasites. *Science* 264:1868–69.
Nishikura, K. 2001. A short primer on RNAi: RNA-directed RNA polymerase acts as a key catalyst. *Cell* 107:415–18.
Seiwert, S.D. 1996. RNA editing hints of a remarkable diversity in gene expression pathways. *Science* 274:1636–37.
Simpson, L. and D.A. Maslov. 1994. RNA editing and the evolution of parasites. *Science* 264:1870–71.
Solner-Webb, B. 1996. Trypanosome RNA editing: Resolved. *Science* 273:1182–83.
Sontheimer, E.J. and R.W. Carthew. 2004. Argonaute journeys into the heart of RISC. *Science* 305:1409–10.

Research Articles

Abraham, J.M., J.E. Feagin, and K. Stuart. 1988. Characterization of cytochrome *c* oxidase III transcripts that are edited only in the 3′ region. *Cell* 55:267–72.
Bagga, S., J. Bracht, S. Hunter, K. Massirer, J. Holtz, R. Eachus, and A.E. Pasquinelli. 2005. Regulation by *let-7* and *lin-4* miRNAs results in target mRNA degradation. *Cell* 122:553–63.
Binder, R., J.A. Horowitz, J.P. Basilion, D.M. Koeller, R.D. Klausner, and J.B. Harford. 1994. Evidence that the pathway of transferrin receptor mRNA degradation involves an endonucleolytic cleavage within the 3′ UTR and does not involve poly(A) tail shortening. *EMBO Journal* 13:1969–80.
Blum, B., N. Bakalara, and L. Simpson. 1990. A model for RNA editing in kinetoplastid mitochondria: "Guide" RNA molecules transcribed from maxicircle DNA provide the edited information. *Cell* 60:189–98.
Casey, J.L., M.W. Hentze, D.M. Koeller, S.W. Caughman, T.A. Rovault, R.D. Klausner, and J.B. Harford. 1988. Iron-responsive elements: Regulatory RNA sequences that control mRNA levels and translation. *Science* 240:924–28.
Casey, J.L., D.M. Koeller, V.C. Ramin, R.D. Klausner, and J.B. Harford. 1989. Iron regulation of transferrin receptor mRNA levels requires iron-responsive elements and a rapid turnover determinant in the 3′ untranslated region of the mRNA. *EMBO Journal* 8:3693–99.
Feagin, J.E., J.M. Abraham, and K. Stuart. 1988. Extensive editing of the cytochrome c oxidase III transcript in *Trypanosoma brucei*. *Cell* 53:413–22.
Fire, A., S. Xu, M.K. Montgomery, S.A. Kostas, S.E. Driwer, and C.C. Mello. 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391:806–11.
Fukagawa, T., M. Nogami, M. Yoshikawa, M. Ikeno, T. Okazaki, Y. Takami, T. Nakayama, and M. Oshimura. 2004. Dicer is essential for formation of the heterochromatin structure in vertebrate cells. *Nature Cell Biology* 6:784–91.
Guerrier-Takada, C., K. Gardiner, T. Marsh, N. Pace, and S. Altman. 1983. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* 35:849–57.
Guyette, W.A., R.J. Matusik, and J.M. Rosen. 1979. Prolactin-mediated transcriptional and post-transcriptional control of casein gene expression. *Cell* 17:1013–23.
Hall, I.M., G.D. Shankaranarayana, K.-i. Noma, N. Ayoub, A. Cohen, and S.I.S. Grewal. 2002. Establishment and maintenance of a heterochromatin domain. *Science* 297:2232–37.
Hammond, S.M., E. Bernstein, D. Beach, and G.J. Hannon. 2000. An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* 404:293–96.
Johnson, P.J., J.M. Kooter, and P. Borst. 1987. Inactivation of transcription by UV irradiation of *T. brucei* provides evidence for a multicistronic transcription unit including a VSG gene. *Cell* 51:273–81.
Kable, M.L., S.D. Seiwart, S. Heidmann, and K. Stuart. 1996. RNA editing: A mechanism for gRNA-specified uridylate insertion into precursor mRNA. *Science* 273:1189–95.
Koeller, D.M., J.L. Casey, M.W. Hentze, E.M. Gerhardt, L.-N.L. Chan , R.D. Klausner, and J.B. Harford. 1989. A cytosolic protein binds to structural elements within the iron regulatory region of the transferrin receptor mRNA. *Proceedings of the National Academy of Sciences USA* 86:3574–78.
Koeller, D.M., J.A. Horowitz, J.L. Casey, R.D. Klausner, and J.B. Harford. 1991. Translation and the stability of mRNAs encoding the transferrin receptor and *c-fos*. *Proceedings of the National Academy of Sciences USA* 88:7778–82.
Lee, R.C., R.L. Feinbaum, and V. Ambros. 1993. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75:843–54.
Li, Z. and M.P. Deutscher. 1994. The role of individual exoribonucleases in processing at the 3′ end of *Escherichia. coli* tRNA precursors. *Journal of Biological Chemistry* 269:6064–71.
Lipardi, C., Q. Wei, and B.M. Paterson. 2001. RNAi as random degradative PCR: siRNA primers convert mRNA into dsRNAs that are degraded to generate new siRNAs. *Cell* 107:297–307.
Liu, J., M.A. Carmell, F.V. Rivas, C.G. Marsden, J.M. Thomson, J.-J. Song, S.M. Hammond, L. Joshua-Tor, and G.J. Hannon. 2004. Argonaute2 is the catalytic engine of mammalian RNAi. *Science* 305:1437–41.
Miller, O.L., Jr., B.R. Beatty, B.A. Hamkalo, and C.A. Thomas, Jr. 1970. Electron microscopic visualization of transcription. *Cold Spring Harbor Symposia on Quantitative Biology.* 35:505–12.
Morris, K.V., S.W.-L. Chan, S.E. Jacobsen, and D.J. Looney. 2004. Small interfering RNA-induced transcriptional gene silencing in human cells. *Science* 305:1289–92.
Müllner, E.W. and L.C. Kühn. 1988. A stem-loop in the 3′ untranslated region mediates iron-dependent regulation of transferrin receptor mRNA stability in the cytoplasm. *Cell* 53:815–25.
Murphy, W.J., K.P. Watkins, and N. Agabian. 1986. Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: Evidence for *trans* splicing. *Cell* 47:517–25.

Olsen, P.H. and V. Ambros. 1999. The *lin-4* regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. *Developmental Biology* 216:671–80.
Owen, D. and L.C. Kühn. 1987. Noncoding 3' sequences of the transferrin receptor gene are required for mRNA regulation by iron. *EMBO Journal* 6:1287–93.
Seiwert, S.D. and K. Stuart. 1994. RNA editing: Transfer of genetic information from gRNA to precursor mRNA in vitro. *Science* 266:114–17.
Sijen, T. J. Fleenor, F. Simmer, K.L. Thijssen, S. Parrish, L. Timmons, R.H.A. Plasterk, and A. Fire. 2001. On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell* 107:465–76.
Song, J.-J., J. Liu, N.H. Tolia, J. Schneiderman, S.K. Smith, R.A. Martienssen, G.J. Hannon, and L. Joshua-Tor. 2003. The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nature Structural Biology* 10:1026–32.
Song, J.-J., S.K. Smith, G.J. Hannon, and L. Joshua-Tor. 2004. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* 305:1434–37.
Volpe, T.A., Kidner, I.M. Hall, G. Teng, S.I.S. Grewal, and R.A. Martienssen. 2002. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* 297:1833–37.
Wang, Q., J. Khillan, P. Gaude, and K. Nishikura. 2000. Requirement of the RNA editing deaminase ADAR1 gene for embryonic erythropoiesis. *Science* 290:1765–8.
Weinberg, R.A., and S. Penman. 1970. Processing of 45S nucleolar RNA. *Journal of Molecular Biology* 47:169–78.
Yekta, S., I.-h. Shih, and D.P. Bartel. 2004. MicroRNA-directed cleavage of *HOXB8* mRNA. *Science* 304:594–96.
Zamore, P.D., T. Tuschl, P.A. Sharp, and D.P. Bartel. 2000. RNAi: Double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* 101:25–33.

第十七章

General References and Reviews

Cech, T.R. 2004. RNA finds a simpler way. *Nature* 428:263–64.
Gottesman, S. 2004. The small RNA regulators of *Escherichia coli:* Roles and mechanisms. *Annual Review of Microbiology* 58:303–28.
Hentze, M.W. 1997. eIF4G: A multipurpose ribosome adapter? *Science* 275:500–1.
Jackson, R.J. 1998. Cinderella factors have a ball. *Nature* 394:829–31.
Kozak, M. 1989. The scanning model for translation: An update. *Journal of Cell Biology* 108:229–41.
Kozak, M. 1991. Structural features in eukaryotic mRNAs that modulate the initiation of translation. *Journal of Biological Chemistry* 266:19867–70.
Lawrence, J.C. and Abraham, R.T. 1997. PHAS/4E-BPs as regulators of mRNA translation and cell proliferation. *Trends in Biochemical Sciences.* 22:345–49.
Proud, C.G. 1994. Turned on by insulin. *Nature* 371:747–48.
Rhoads, R.E. 1993. Regulation of eukaryotic protein synthesis by initiation factors. *Journal of Biological Chemistry* 268:3017–20.
Richter, J.D. and W.E. Theurkauf. 2001. The message is in the translation. *Science* 293:60–62.
Roll-Mecak, A., B.-S. Shin, T.E. Dever, and S.K. Burley. 2001. Engaging the ribosome: Universal IFs of translation. *Trends in Biochemical Sciences* 26:705–9.
Sachs, A.B. 1997. Starting at the beginning, middle, and end: Translation initiation in eukaryotes. *Cell* 89:831–38.
Thach, R.E. 1992. Cap recap: The involvement of eIF4F in regulating gene expression. *Cell* 68:177–80.

Research Articles

Aziz, N. and H.N. Munro. 1987. Iron regulates ferritin mRNA translation through a segment of its 5'-untranslated region. *Proceedings of the National Academy of Sciences USA* 84:8478–82.
Brown, L. and T. Elliott. 1997. Mutations that increase expression of the *rpoS* gene and decrease its dependence on *hfq* function in *Salmonella typhimurium. Journal of Bacteriology* 179:656–62.
Cigan, A.M., L. Feng, and T.F. Donahue. 1988. $tRNA_i^{Met}$ functions in directing the scanning ribosome to the start site of translation. *Science* 242:93–96.
Dubnoff, J.S., A.H. Lockwood, and U. Maitra. 1972. Studies on the role of guanosine triphosphate in polypeptide chain initiation in *Escherichia coli. Journal of Biological Chemistry* 247:2884–94.
Edery, I., M. Hümbelin, A. Darveau, K.A.W. Lee, S. Milburn, J.W.B. Hershey, H. Trachsel, and N. Sonenberg. 1983. Involvement of eukaryotic initiation factor 4A in the cap recognition process. *Journal of Biological Chemistry* 258:11398–403.
Fakunding, J.L. and J.W.B. Hershey. 1973. The interaction of radioactive initiation factor IF2 with ribosomes during initiation of protein synthesis. *Journal of Biological Chemistry* 248:4206–12.
Guthrie, C. and M. Nomura. 1968. Initiation of protein synthesis: A critical test of the 30S subunit model. *Nature* 219:232–35.
Hui, A. and H.A. De Boer. 1987. Specialized ribosome system: Preferential translation of a single mRNA species by a subpopulation of mutated ribosomes in *Escherichia coli. Proceedings of the National Academy of Sciences USA* 84:4762–66
Kaempfer, R.O.R., M. Meselson, and H.J. Raskas. 1968. Cyclic dissociation into stable subunits and reformation of ribosomes during bacterial growth. *Journal of Molecular Biology* 31:277–89.
Kozak, M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* 44:283–92.
Kozak, M. 1989. Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs. *Molecular and Cellular Biology* 9:5134–42.
Lin, T.-A. 1994. PHAS-I as a link between mitogen-activated protein kinase and translation initiation. *Science* 266:653–56.
Min Jou, W., G. Haegeman, M. Ysebaert, and W. Fiers. 1972. Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. *Nature* 237:82–88.
Morita, M.T., Y. Tanaka, T.S. Kodama, Y. Kyogoku, K. Yanagi, and T. Yura. 1999. Translational induction of heat shock transcription factor σ^{32}. Evidence for a built-in RNA thermosensor. *Genes and Development* 13:655–65.
Noll, M. and H. Noll. 1972. Mechanism and control of initiation in the translation of R17 RNA. *Nature New Biology* 238:225–28.
Pause, A., G.J. Belsham, A.-C. Gingras, O. Donzé, T.-A. Lin, J.C. Lawrence, and N. Sonenberg. 1994. Insulin-dependent stimulation of protein synthesis by phosphorylation of a regulator of 5'-cap function. *Nature* 371:762–67.
Pause, A. and N. Sonenberg. 1992. Mutational analysis of a DEAD box RNA helicase: The mammalian translation initiation factor eIF4A. *EMBO Journal* 11:2643–54.
Pestova, T.V., S.I. Borukhov, and C.V.T. Hellen. 1998. Eukaryotic ribosomes require initiation factors 1 and 1A to locate initiation codons. *Nature* 394:854–59.
Pestova, T.V., I.B. Lomakin, J.H. Lee, S.K. Choi, T.E. Dever, and C.U.T. Hellen. 2000. The joining of ribosomal subunits in eukaryotes requires eIF5B. *Nature* 403:332–35.
Pillai, R.S., S.N. Bhattacharyya, C.G. Artus, T. Zoller, N. Cougot, E. Basyuk, E. Bertrand, and W. Filipowicz. 2005. Inhibition of translational initiation by Let-7 microRNA in human cells. *Science* 309:1573–76.
Sabol, S. and S. Ochoa. 1971. Ribosomal binding of labeled initiation factor F_3. *Nature New Biology* 234:233–36.

Sabol, S., M.A.G. Sillero, K. Iwasaki, and S. Ochoa. 1970. Purification and properties of initiation factor F_3. *Nature* 228:1269–75.
Siridechadilok, B., C.S. Fraser, R.J. Hall, J.A. Doudna, and E. Nogales. 2005. Structural roles for human translation factor eIF3 in initiation of protein synthesis. *Science* 310:1513–15.
Sonenberg, N., M.A. Morgan, W.C. Merrick, and A.J. Shatkin. 1978. A polypeptide in eukaryotic initiation factors that crosslinks specifically to the 5′-terminal cap in mRNA. *Proceedings of the National Academy of Sciences USA* 75:4843–47.
Sonenberg, N., H. Trachsel, S. Hecht, and A.J. Shatkin. 1980. Differential stimulation of capped mRNA translation in vitro by cap binding protein. *Nature* 285:331–33.
Steitz, J.A. and K. Jakes. 1975. How ribosomes select initiator regions in mRNA: Base pair formation between the 3′-terminus of 16S rRNA and the mRNA during initiation of protein synthesis in *Escherichia coli*. *Proceedings of the National Academy of Sciences USA* 72:4734–38.
Wahba, A.J., K. Iwasaki, M. J. Miller, S. Sabol, M.A.G. Sillero, and C. Vasquez. 1969. Initiation of protein synthesis in *Escherichia coli*, II. Role of the initiation factors in polypeptide synthesis. *Cold Spring Harbor Symposia* 34:291–99.
Winkler, W., A. Nahvi, and R.R. Breaker. 2002. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* 419:952–56.

第十八章

General References and Reviews

Horwich, A. 2004. Sight at the end of the tunnel. *Nature* 431:520–22.
Kaji, A., M.C. Kiel, G. Hirokawa, A.R. Muto, Y. Inokuchi, and H. Kaji. 2001. The fourth step of protein synthesis: Disassembly of the posttermination complex is catalyzed by elongation factor G and ribosome recycling factor, a near-perfect mimic of tRNA. *Cold Spring Harbor Symposia on Quantitative Biology* 66:515–29.
Kaziro, Y. 1978. The role of guanosine 5′-triphosphate in polypeptide chain elongation. *Biochimica et Biophysica Acta* 505:95–127.
Khorana, H.G. 1968. Synthesis in the study of nucleic acids. *Biochemistry Journal* 109:709–25.
Maquat, L.E. 2002. Skiing toward nonstop mRNA decay. *Science* 295:2221–22.
Moore, M.J. 2002. No end to nonsense. *Science* 298:370–71
Moore, S.D., K.E. McGinness, and R.T. Sauer. 2003. A glimpse into tmRNA-mediated ribosome rescue, *Science* 300:72–73.
Nakamura, Y., K. Ito, and L.A. Isaksson. 1996. Emerging understanding of translation termination. *Cell* 87:147–50.
Nakamura, Y., K. Ito, and M. Ehrenberg. 2000. Mimicry grasps reality in translation termination. *Cell* 101:349–52.
Nierhaus, K.H. 1996. An elongation factor turn-on. *Nature* 379:491–92.
Ramakrishnan, V. 2002. Ribosome structure and the mechanism of translation. *Cell* 108:557–72.
Schimmel, P. and K. Beebe. 2004. Genetic code seizes pyrrolysine. *Nature* 431:257–58.
Thompson, R.C. 1988. EFTu provides an internal kinetic standard for translational accuracy. *Trends in Biochemical Sciences* 13:91–93.
Tuite, M.F. and I. Stansfield. 1994. Knowing when to stop. *Nature* 372:614–15.

Research Articles

Benzer, S. and S.P. Champe. 1962. A change from nonsense to sense in the genetic code. *Proceedings of the National Academy of Sciences USA* 48:1114–21.
Brenner, S., A.O.W. Stretton, and S. Kaplan. 1965. Genetic code: The "nonsense" triplets for chain termination and their suppression. *Nature* 206:994–98.
Bretscher, M.S., and K.A. Marcker. 1966. Peptidyl-sRibonucleic acid and amino-acyl-sRibonucleic acid binding sites on ribosomes. *Nature* 211:380–84.
Dintzis, H.M. 1961. Assembly of the peptide chains of hemoglobin. *Proceedings of the National Academy of Sciences USA* 47:247–61.
Ferbitz, L., T. Maier, H. Patzelt, B. Bukau, E. Deuerling, and N. Ban. 2004. Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature* 431:590–96.
Fredrick, K. and H.F. Noller. 2003. Catalysis of ribosomal translocation by sparsomycin. *Science* 300:1159–62.
Frovola, L., X. Le Goff, H.H. Rasmussen, S. Cheperegin, G. Drugeon, M. Kress, I. Arman, A.-L. Haenni, J.E. Celis, M. Philippe, J. Justesen, and L. Kisselev. 1994. A highly conserved eukaryotic protein family possessing properties of polypeptide chain release factor. *Nature* 372:701–3.
Goodman, H.M., J. Abelson, A. Landy, S. Brenner, and J.D. Smith. 1968. Amber suppression: A nucleotide change in the anticodon of a tyrosine transfer RNA. *Nature* 217:1019–24.
Gordon, J. 1968. A stepwise reaction yielding a complex between a supernatant fraction from *E. coli*, guanosine 5′-triphosphate, and aminoacyl-sRNA. *Proceedings of the National Academy of Sciences USA* 59:179–83.
Haenni, A.-L. and J. Lucas-Lenard. 1968. Stepwise synthesis of a tripeptide. *Proceedings of the National Academy of Sciences USA* 61:1363–69.
Inoue-Yokosawa, N., C. Ishikawa, and Y. Kaziro. 1974. The role of guanosine triphosphate in translocation reaction catalyzed by elongation factor G. *Journal of Biological Chemistry* 249:4321–23.
Ito, K., M. Uno, and Y. Nakamura. 2000. A tripeptide "anticodon" deciphers stop codons in messenger RNA. *Nature* 403:680–84.
Khaitovich, P., A.S. Mankin, R. Green, L. Lancaster, and H.F. Noller. 1999. Characterization of functionally active subribosomal particles from *Thermus aquaticus*. *Proceedings of the National Academy of Sciences USA* 96:85–90.
Lancaster, L., M.C. Kiel, A. Kaji, and H.F. Noller. 2002. Orientation of ribosome recycling factor in the robosome from directed hydroxyl radical probing. *Cell* 111:129–40.
Last, J.A., W.M. Stanley, Jr., M. Salas, M.B. Hille, A.J. Wahba, and S. Ochoa. 1967. Translation of the genetic message, IV. UAA as a chain termination codon. *Proceedings of the National Academy of Sciences USA* 57:1062–67.
Miller, D.L. and H. Weissbach. 1970. Interactions between the elongation factors: The displacement of GDP from the Tu-GDP complex by factor Ts. *Biochemical and Biophysical Research Communications* 38:1016–22.
Nirenberg, M. and P. Leder. 1964. RNA codewords and protein synthesis: The effect of trinucleotides upon binding of sRNA to ribosomes. *Science* 145:1399–1407.
Nissen, P., M. Kjeldgaard, S. Thirup, G. Polekhina, L. Reshetnikova, B.F.C. Clark, and J. Nyborg. 1995. Crystal structure of the ternary complex of Phe-tRNA^{Phe}, EF-Tu, and a GTP analog. *Science* 270:1464–71.
Noller, H.F., V. Hoffarth, and L. Zimniak. 1992. Unusual resistance of peptidyl transferase to protein extraction procedures. *Science* 256:1416–19.
Ravel, J.M. 1967. Demonstration of a guanine triphosphate-dependent enzymatic binding of aminoacyl-ribonucleic acid to *Escherichia coli* ribosomes. *Proceedings of the National Academy of Sciences USA* 57:1811–16.
Ravel, J.M., R.L. Shorey, and W. Shire. 1968. The composition of the active intermediate in the transfer of aminoacyl-RNA to ribosomes. *Biochemical and Biophysical Research Communications* 32:9–14.
Rheinberger, H.-J., H. Sternbach, and K.H. Nierhaus. 1981. Three tRNA binding sites on *Escherichia coli* ribosomes. *Proceedings of the National Academy of Sciences USA* 78:5310–14.
Rodnina, M.V., A. Savelsbergh, V.I. Katunin, and W.

Wintermeyer. 1997. Hydrolysis of GTP by elongation factor G drives tRNA movement on the ribosome. *Nature* 385:37–41.
Sarabhai, A.S., A.O.W. Stretton, S. Brenner, and A. Bolle. 1964. Co-linearity of the gene with the polypeptide chain. *Nature* 201:13–17.
Scolnick, E., R. Tompkins, T. Caskey, and M. Nirenberg. 1968. Release factors differing in specificity for terminator codons. *Proceedings of the National Academy of Sciences USA* 61:768–74.
Song, H., P. Mugnier, A.K. Das, H.M. Webb, D.R. Evans, M. F. Tuite, B.A. Hemmings, and D. Barford. 2000. The crystal structure of human eukaryotic release factor eRF1—Mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. *Cell* 100:311–21.
Thach, S.S. and R.E. Thach. 1971. Translocation of messenger RNA and "accommodation" of fMet-tRNA. *Proceedings of the National Academy of Sciences USA* 68:1791–95.
Traut, R.R. and R.E. Monro. 1964. The puromycin reaction and its relation to protein synthesis. *Journal of Molecular Biology* 10:63–72.
Valle, M., R. Gillet, S. Kaur, A. Henne, V. Ramakrishnan, and J. Frank. 2003 Visualizing tmRNA entry into a stalled ribosome. *Science* 300:127–30.
Weigert, M.G. and A. Garen. 1965. Base composition of nonsense codons in *E. coli*. *Nature* 206:992–94.
Weissbach, H., D.L. Miller, and J. Hachmann. 1970. Studies on the role of factor Ts in polypeptide synthesis. *Archives of Biochemistry and Biophysics* 137:262–69.
Zhouravleva, G., L. Frolova, X. Le Goff, R. Le Guellec, S. Inge-Vechtomov, L. Kisselev, and M. Philippe. 1995. Termination of translation in eukaryotes is governed by two interacting polypeptide chain release factors, eRF1 and eRF3. *EMBO Journal* 14:4065–72.

第十九章

General References and Reviews

Cech, T.R. 2000. The ribosome is a ribozyme. *Science* 289:878–79.
Dahlberg, A.E. 2001. The ribosome in action. *Science* 292:868–69.
Fersht, A.R. 1998. Sieves in sequence. *Science* 280:541.
Hill, W.E., P.B. Moore, A. Dahlberg, D. Schlessinger, R.A. Garrett, and J.R. Warner, eds. 1990. *The Ribosome: Structure, Function and Evolution*. Washington, D.C.: American Society for Microbiology.
Moore, P.B. 2005. A ribosomal coup: *E. coli* at last! *Science* 310:793–95.
Noller, H.F. 1990. Structure of rRNA and its functional interactions in translation. In Hill, W.E., et al., eds. *The Ribosome: Structure, Function and Evolution*. Washington, D.C.: American Society for Microbiology, chapter 3, pp. 73–92.
Pennisi, E. 2001. Ribosome's inner workings come into sharper view. *Science* 291:2526–27.
Saks, M.E., J.R. Sampson, and J.N. Abelson. 1994. The transfer RNA identity problem: A search for rules. *Science* 263:191–97.
Waldrop, M.M. 1990. The structure of the "second genetic code." *Science* 246:1122.
Zimmermann, R.A. 1995. Ins and outs of the ribosome. *Nature* 376:391–92.

Research Articles

Ban, N., P. Nissen, J. Hansen, P.B. Moore, and T.A. Steitz. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* 289:905–20.
Carter, A.P., W.M. Clemons, Jr., D.E. Brodersen, R.J. Morgan-Warren, T. Hartsch, B.T. Wimberly, and V. Ramakrishnan. 2000. Crystal structure of an initiation factor bound to the 30S ribosomal subunit. *Science* 291:498–501.
Carter, A.P., W.M. Clemons, Jr., D.E. Brodersen, R.J. Morgan-Warren, T. Hartsch, B.T. Wimberly, and V. Ramakrishnan. 2000. Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature* 407:340–48.
Held, W.A., B. Ballou, S. Mizushima, and M. Nomura. 1974. Assembly mapping of 30S ribosomal proteins from *Escherichia coli*: Further studies. *Journal of Biological Chemistry* 249:3103–11.
Hoagland, M.B., M.L. Stephenson, J.F. Scott, L.I. Hecht, and P.C. Zamecnik. 1958. A soluble ribonucleic acid intermediate in protein synthesis. *Journal of Biological Chemistry* 231:241–57.
Holley, R.W., J. Apgar, G.A. Everett, J.T. Madison, M. Marquisee, S.H. Merrill, J.R. Penswick, and A. Zamir. 1965. Structure of a ribonucleic acid. *Science* 147:1462–65.
Kaltschmidt, E. and H.G. Wittmann. 1970. Ribosomal proteins XII: Number of proteins in small and large ribosomal subunits of *Escherichia coli* as determined by two-dimensional gel electrophoresis. *Proceedings of the National Academy of Sciences USA* 67:1276–82.
Kim, S.H., F.L. Suddath, G.J. Quigley, A. McPherson, J.L. Sussman, A.H.J. Wang, N. C. Seeman, and A. Rich. 1974. Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. *Science* 185:435–40.
Lake, J.A. 1976. Ribosome structure determined by electron microscopy of *Escherichia coli* small subunits, large subunits and monomeric ribosomes. *Journal of Molecular Biology* 105:131–59.
Miller, O., B.A. Hamkalo, and C.A. Thomas, Jr. 1970. Visualization of bacterial genes in action. *Science* 169:392–95.
Mizushima, S. and M. Nomura. 1970. Assembly mapping of 30S ribosomal proteins from *E. coli*. *Nature* 226:1214–18.
Muth, G.W., L. Ortoleva-Donnelly, and S.A. Strobel. 2000. A single adenosine with a neutral pK_a in the ribosomal peptidyl transferase center. *Science* 289:947–50.
Nissen, P., J. Hansen, N. Ban, P.B. Moore, and T.A. Steitz. 2000. The structural basis of ribosome activity in peptide bond synthesis. *Science* 289:920–30.
Nureki, O., D.G. Vassylyev, M. Tateno, A. Shimada, T. Nakama, S. Fukai, M. Konno, T.L. Henrickson. P. Schimmel, and S. Yokoyama. 1998. Enzyme structure with two catalytic sites for double-sieve selection of substrates. *Science* 280:578–82.
Ogle, J.M., D.E Brodersen, W.M.Clemons Jr., M.J. Tarry, A.P. Carter, and V. Ramakrishnan. 2001. Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* 292:897–902.
Pallanck, L. and L.H. Schulman. 1991. Anticodon-dependent aminoacylation on a noncognate tRNA with isoleucine, valine, and phenylalanine in vivo. *Proceedings of the National Academy of Sciences USA* 88:3872–76.
Polacek, N., M. Gaynor, A. Yassin, and A.S Mankin. 2001. Ribosomal peptidyl transferase can withstand mutations at the putative catalytic nucleotide. *Nature* 411:498–501.
Quigley, G.J. and A. Rich. 1976. Structural domains of transfer RNA molecules. *Science* 194:796–806.
Rould, M.A., J.J. Perona, D. Söll, and T.A. Steitz. 1989. Structure of *E. coli* glutaminyl-tRNA synthetase complexed with $tRNA^{Gln}$ and ATP at 2.8 Å resolution. *Science* 246:1135–42.
Ruff, M., S. Krishnaswamy, M. Boeglin, A. Poterszman, A. Mitschler, A. Podjarny, B. Rees, J.C. Thierry, and D. Moras. 1991. Class II aminoacyl transfer RNA synthetases: Crystal structure of yeast aspartyl-tRNA synthetase complexed with $tRNA^{Asp}$. *Science* 252:1682–89.
Schluenzen, F., A. Tocilj, R. Zarivach, J. Harms, M. Gluehmann, D. Janell, A. Bashan, H. Bartels, I. Agmon, F. Franceschi, and A. Yonath. 2000. Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. *Cell* 102:615–23.
Schulman, L.H. and H. Pelka. 1983. Anticodon loop size and sequence requirements for recognition of formylmethionine

tRNA by methionyl-tRNA synthetase. *Proceedings of the National Academy of Sciences USA* 80:6755–59.
Schuwirth, B.S., M.A. Borovinskaya, C.W. Hau, W. Zhang, A. Vila-Sanjurjo, J.M. Holton, and J.H. Doundna Cate. 2005. Structures of the bacterial ribosome at 3.5 Å resolution. *Science* 310:827–34.
Stern, S., B. Weiser, and H.F. Noller. 1988. Model for the three-dimensional folding of 16S ribosomal RNA. *Journal of Molecular Biology* 204:447–81.
Weinger, J.S., K.M. Parnell, S. Dorner, R. Green, and S.A. Strobel. 2004. Substrate-assisted catalysis of peptide bond formation by the ribosome. *Nature Structural and Molecular Biology* 11:1101–06.
Wimberly, B.T., D.E. Brodersen, W.M. Clemons Jr., R.J. Morgan-Warren, A.P. Carter, C. Vonrhein, T. Hartsch, and V. Ramakrishnan. 2000. Structure of the 30S ribosomal subunit. *Nature* 407:327–39.
Yusupov, M.M., G. Zh. Yusupova, A. Baucom, K. Lieberman, T.N. Earnest, J.H.D. Cate, and H.F Noller. 2001. Crystal structure of the ribosome at 5.5 Å resolution. *Science* 292:883–96.

第二十章

General References and Reviews

Cairns, B.R. 2004. Around the world of DNA damage INO80 days. *Cell* 119:733–34.
Chu, G. 1997. Double strand break repair. *Journal of Biological Chemistry.* 272:24097–100.
Citterio, E., W. Vermeulen, and J.H.J. Hoeijmakers. 2000. Transcriptional healing. *Cell* 101:447–50.
David, S.S. 2005. DNA search and rescue. *Nature* 434:569–70.
de Latt, W.L., N.G.J. Jaspers, and J.H.J. Hoeijmakers. 1999. Molecular mechanism of nucleotide excision repair. *Genes and Development* 13:768–85.
Friedberg, E.C., R. Wagner, and M. Radman. 2002. Specialized DNA polymerases, cellular survival, and the genesis of mutations. *Science* 296:1627–30.
Herendeen, D.R. and T.J. Kelly. 1996. DNA polymerase III: Running rings around the fork. *Cell* 84:5–8.
Jiricny, J. 2002. An APE that proofreads. *Nature* 415:593–94.
Joyce, C.M. and T.A. Steitz. 1987. DNA polymerase I: From crystal structure to function via genetics. *Trends in Biochemical Sciences* 12:288–92.
Kornberg, A. and T. Baker. 1992. *DNA Replication.* New York: W.H. Freeman and Company.
Lindahl, T. 2004. Molecular biology: Ensuring error-free DNA repair. *Nature* 427:598.
Lindahl, T. and R.D. Wood. 1999. Quality control by DNA repair. *Science* 286:1897–1905.
Maxwell, A. 1996. Protein gates in DNA topoisomerase II. *Nature Structural Biology.* 3:109–12.
Radman, M. and R. Wagner. 1988. The high fidelity of DNA replication. *Scientific American* (August):40–46.
Sharma, A. and A. Mondragón. 1995. DNA topoisomerases. *Current Opinion in Structural Biology* 5:39–47.
Wood, R.D. 1997. Nucleotide excision repair in mammalian cells. *Journal of Biological Chemistry* 272:23465–68.
Wood, R.D. 1999. Variants on a theme. *Nature* 399:639–70.

Research Articles

Bagg, A., C.J. Kenyon, and G.C. Walker. 1981. Inducibility of a gene product required for UV and chemical mutagenesis in *Escherichia coli. Proceedings of the National Academy of Sciences USA* 78:5749–53.
Banerjee, A., W. Yang, M. Karplus, and G.L. Verdine. 2005. Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA. *Nature* 434:612–18.
Berger, J.M., S.J. Gamblin, S.C. Harrison, and J.C. Wang. 1996. Structure and mechanism of DNA topoisomerase II. *Nature* 379:225–32.
Cairns, J. 1963. The chromosome of *Escherichia coli. Cold Spring Harbor Symposia on Quantitative Biology* 28:43–46.
Chou, K.-M. and Y.-C. Cheng. 2002. An exonucleolytic activity of human apurinic/apyrimidinic endonuclease on 3′ mispaired DNA. *Nature* 415:655–59.
Curtis, M.J. and B. Alberts. 1976. Studies on the structure of intracellular bacteriophage T4 DNA. *Journal of Molecular Biology* 102:793–816.
Drapkin, R., J.T. Reardon, A. Ansari, J.-C. Huang, L. Zawel, K. Ahn, A. Sancar, and D. Reinberg. 1994. Dual role of TFIIH in DNA excision repair and in transcription by RNA polymerase II. *Nature* 368:769–72.
Eom, S.H., T. Wang, and T.A. Steitz. 1996. Structure of *Taq* polymerase with DNA at the active site. *Nature* 382: 278–281.
Gefter, M.L., Y. Hirota, T. Kornberg, J.A. Wechster, and C. Barnoux. 1971. Analysis of DNA polymerases II and III in mutants of *Escherichia coli* thermosensitive for DNA synthesis. *Proceedings of the National Academy of Sciences USA* 68:3150–53.
Gellert, M., K. Mizuuchi, M.H. O'Dea, and H.A. Nash. 1976. DNA gyrase: An enzyme that introduces superhelical turns into DNA. *Proceedings of the National Academy of Sciences USA* 73:3872–76.
Gyurasits, E.B. and R.J. Wake. 1973. Bidirectional chromosome replication in *Bacillus subtilis. Journal of Molecular Biology* 73:55–63.
Hirota, G.H., A. Ryter, and F. Jacob. 1968. Thermosensitive mutants in *E. coli* affected in the processes of DNA synthesis and cellular division. *Cold Spring Harbor Symposia on Quantitative Biology* 33:677–93.
Huberman, J.A., A. Kornberg, and B.M. Alberts. 1971. Stimulation of T4 bacteriophage DNA polymerase by the protein product of T4 gene 32. *Journal of Molecular Biology* 62:39–52.
Kitani, T., K.-Y. Yoda, T. Ogawa, and T. Okazaki. 1985. Evidence that discontinuous DNA replication in *Escherichia coli* is primed by approximately 10 to 12 residues of RNA starting with a purine. *Journal of Molecular Biology* 184:45–52.
LeBowitz, J.H. and R. McMacken. 1986. The *Escherichia coli* dnaB replication protein is a DNA helicase. *Journal of Biological Chemistry* 261:4738–48.
Maki, H. and A. Kornberg. 1985. The polymerase subunit of DNA polymerase III of *Escherichia coli. Journal of Biological Chemistry* 260:12987–92.
Masutani, C., R. Kusumoto, A. Yamada, M. Yuasa, M. Araki, T. Nogimori, M. Yokoi, T. Eki, S. Iwai, and F. Hanaoka. 2000. Xeroderma pigmentosum variant: From a human genetic disorder to a novel DNA polymerase. *Cold Spring Harbor Symposia on Quantitative Biology.* 65:71–80.
Masutani, C., R. Kusumoto, A. Yamada, N. Dohmae, M. Yokoi, M. Yuasa, M. Araki, S. Iwai, K. Takio, and F. Hanaoka. 1999. The *XPV* (xeroderma pigmentosum variant) gene encodes human DNA polymerase η. *Nature* 399: 700–04.
Matsuda, T., K. Bebenek, C. Masutani, F. Hanaoka, and T.A. Kunkel. 2000. Low fidelity DNA synthesis by human DNA polymerase-η. *Nature* 404:1011–13.
Meselson, M. and F. Stahl. 1958. The replication of DNA in *Escherichia coli. Proceedings of the National Academy of Sciences USA* 44:671–82.
Okazaki, R., T. Okazaki, K. Sakabe, K. Sugimoto, R. Kainuma, A. Sugino, and N. Iwatsuki. 1968. In vivo mechanism of DNA chain growth. *Cold Spring Harbor Symposia on Quantitative Biology* 3:129–43.
Scheuermann, R.H. and H. Echols. 1984. A separate editing exonuclease for DNA replication: The ε subunit of *Escherichia coli* DNA polymerase III holoenzyme. *Proceedings of the National Academy of Sciences USA* 81:7747–57.

Sugasawa, K., J.M.Y. Ng, C. Masutani, S. Iwai, P.J. van der Spek, A.P.M. Eker, F. Hanaoka, D. Bootsma, and J.H.J. Hoeijmakers. 1998. Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. *Molecular Cell* 2:223–32.
Tse, Y.-C., K. Kirkegaard, and J.C. Wang. 1980. Covalent bonds between protein and DNA. *Journal of Biological Chemistry* 255:5560–65.
Wakasugi, M. and A. Sancar. 1999. Order of assembly of human DNA repair excision nuclease. *Journal of Biological Chemistry* 274:18759–68.

第二十一章

General References and Reviews

Baker, T.A. 1995. Replication arrest. *Cell* 80:521–24.
Blackburn, E.H. 1990. Telomeres: Structure and synthesis. *Journal of Biological Chemistry* 265:5919–21.
Blackburn, E.H. 1994. Telomeres: No end in sight. *Cell* 77:621–23.
Cech, T. R. 2004. Beginning to understand the end of the chromosome. *Cell* 116:273–79.
de Lange, T. 2001. Telomere capping—one strand fits all. *Science* 292:1075–76.
Ellison, V. and B. Stillman. 2001. Opening of the clamp: An intimate view of an ATP-driven biological machine. *Cell* 106:655–60.
Greider, C.W. 1999. Telomeres do D-loop-T-loop. *Cell* 97:419–22.
Herendeen, D.R. and T.J. Kelly. 1996. DNA polymerase III: Running rings around the fork. *Cell* 84:5–8.
Kornberg, A. and T.A. Baker. 1992. *DNA Replication,* 2nd ed. New York: W.H. Freeman.
Marx, J. 1994. DNA repair comes into its own. *Science* 266:728–30.
Marx, J. 1995. How DNA replication originates. *Science* 270:1585–86.
Marx, J. 2002. Chromosome end game draws a crowd. *Science* 295:2348–51.
Newlon, C.S. 1993. Two jobs for the origin replication complex. *Science* 262:1830–31.
Stillman, B. 1994. Smart machines at the DNA replication fork. *Cell* 78:725–28.
Wang, J.C. 1991. DNA topoisomerases: Why so many? *Journal of Biological Chemistry* 266:6659–62.
West, S.C. 1996. DNA helicases: New breeds of translocating motors and molecular pumps. *Cell* 86:177–80.
Zakian, V.A. 1995. Telomeres: Beginning to understand the end. *Science* 270:1601–6.

Research Articles

Adams, D.E., E.M. Shekhtman, E.L. Zechiedrich, M.B. Schmid, and N.R. Cozzarelli. 1992. The role of topoisomerase IV in partitioning bacterial replicons and the structure of catenated intermediates in DNA replication. *Cell* 71:277–88.
Arai, K. and A. Kornberg. 1979. A general priming system employing only *dnaB* protein and primase for DNA replication. *Proceedings of the National Academy of Sciences USA* 76:4309–13.
Arai, K., R. Low, J. Kobori, J. Shlomai, and A. Kornberg. 1981. Mechanism of *dnaB* protein action V. Association of *dnaB* protein, protein n′, and other prepriming proteins in the primosome of DNA replication. *Journal of Biological Chemistry* 256:5273–80.
Baumann, P. and T. Cech. 2001. Pot 1, the putative telomere end-binding protein in fission yeast and humans. *Science* 292:1171–75.
Blackburn, E.H. 1990. Functional evidence for an RNA template in telomerase. *Science* 247:546–52.
Blackburn, E.H. 2001. Switching and signaling at the telomere. *Cell* 106:661–73.
Bouché, J.-P., L. Rowen, and A. Kornberg. 1978. The RNA primer synthesized by primase to initiate phage G4 DNA replication. *Journal of Biological Chemistry* 253:765–69.
Brewer, B.J. and W.L. Fangman. 1987. The localization of replication origins on ARS plasmids in *S. cerevisiae. Cell* 51:463–71.
Burgers, P.M.J. and A. Kornberg. 1982. ATP activation of DNA polymerase III holoenzyme of *Escherichia coli. Journal of Biological Chemistry* 257:11468–73.
Deb, S., A.L. DeLucia, A. Koff, S. Tsui, and P. Tegtmeyer. 1986. The adenine-thymine domain of the simian virus 40 core antigen directs DNA bending and coordinately regulates DNA replication. *Molecular and Cellular Biology* 6:4578–84.
Fareed, G.C., C.F. Garon, and N.P. Solzman. 1972. Origin and direction of SV40 deoxyribonucleic acid replication. *Journal of Virology* 10:484–91.
Greider, C.W. and E.H. Blackburn. 1985. Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts. *Cell* 43:405–13.
Greider, C.W. and E.H. Blackburn. 1989. A telomeric sequence in the RNA of *Tetrahymena* telomerase required for telomere repeat synthesis. *Nature* 337:331–37.
Griffith, J.D., L. Comeau, S. Rosenfield, R.M. Stansel, A. Bianchi, H. Moss, and T. de Lange. 1999. Mammalian telomeres end in a large duplex loop. *Cell* 97:503–19.
Jeruzalmi, D., M. O'Donnell, and J. Kuriyan. 2001. Crystal structure of the processivity clamp loader gamma (γ) complex of *E. coli* DNA polymerase III. *Cell* 106:429–41.
Jeruzalmi, D., O. Yurieva, Y. Zhao, M. Young, J. Stewart, M. Hingorani, M. O'Donnell, and J. Kuriyan. 2001. Mechanism of processivity clamp opening by the delta subunit wrench of the clamp loader complex of *E. coli* DNA polymerase III. *Cell* 106:417–28.
Kong, X.-P., R. Onrust, M. O'Donnell, and J. Kuriyan. 1992. Three-dimensional structure of the β subunit of *E. coli* DNA polymerase III holoenzyme: A sliding DNA clamp. *Cell* 69:425–37.
Krishna, T.S.R., X.-P. Kong, S. Gary, P.M. Burgers, and J. Kuriyan. 1994. Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell* 79:1233–43.
Marahrens, Y. and B. Stillman. 1992. A yeast chromosomal origin of DNA replication defined by multiple functional elements. *Science* 255:817–23.
Mok, M. and K.J. Marians. 1987. The *Escherichia coli* preprimosome and DNA B helicase can form replication forks that move at the same rate. *Journal of Biological Chemistry* 262:16644–54.
Naktinis, V., J. Turner, and M. O'Donnell. 1996. A molecular switch in a replication machine defined by an internal competition for protein rings. *Cell* 84:137–45.
Stukenberg, P.T., P.S. Studwell-Vaughan, and M. O'Donnell. 1991. Mechanism of the sliding β-clamp of DNA polymerase III holoenzyme. *Journal of Biological Chemistry* 266:11328–34.
Stukenberg, P.T., J. Turner, and M. O'Donnell. 1994. An explanation for lagging strand replication: Polymerase hopping among DNA sliding clamps. *Cell* 78:877–87.

第二十二章

General References and Reviews

Amundsen, S.K. and G.R. Smith. 2003. Interchangeable parts of the *Escherichia coli* recombination machinery. *Cell* 112:741–44.
Fincham, J.R.S. and P. Oliver. 1989. Initiation of recombination. *Nature* 338:14–15.

McEntee, K. 1992. RecA: From locus to lattice. *Nature* 355:302–3.
Meselson, M. and C.M. Radding. 1975. A general model for genetic recombination. *Proceedings of the National Academy of Sciences USA* 72:358–61.
Roeder, G.S. 1997. Meiotic chromosomes: It takes two to tango. *Genes and Development* 11:2600–21.
Smith, G.R. 1991. Conjugational recombination in *E. coli:* Myths and mechanisms. *Cell* 64:19–27.
West, S.C. 1998. RuvA gets x-rayed on Holliday. *Cell* 94:699–701.

Research Articles

Cao, L., E. Alani, and N. Kleckner. 1990. A pathway for generation and processing of double-strand breaks during meiotic recombination in *S. cerevisiae*. *Cell* 61:1089–1101.
DasGupta, C., T. Shibata, R.P. Cunningham, and C.M. Radding. 1980. The topology of homologous pairing promoted by RecA protein. *Cell* 22:437–46.
Dunderdale, H.J., F.E. Benson, C.A. Parsons, G.J. Sharples, R.G. Lloyd, and S.C. West. 1991. Formation and resolution of recombination intermediates by *E. coli* RecA and RuvC proteins. *Nature* 354:506–10.
Eggleston, A.K., A.H. Mitchell, and S.C. West. 1997. In vitro reconstitution of the late steps of genetic recombination in *E. coli*. *Cell* 89:607–17.
Honigberg, S.M., D.K. Gonda, J. Flory, and C.M. Radding. 1985. The pairing activity of stable nucleoprotein filaments made from RecA protein, single-stranded DNA, and adenosine 5′-(γ-thio)triphosphate. *Journal of Biological Chemistry* 260:11845–51.
Keeney, S, C.N. Giroux, and N. Kleckner. 1997. Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell* 88:375–84.
Neale, M.J., J. Pan, and S. Keeney. 2005. Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. *Nature* 436:1053–57.
Parsons, C.A. and S.C. West. 1993. Formation of a RuvAB–Holliday junction complex in vitro. *Journal of Molecular Biology* 232:397–405.
Ponticelli, A.S., D.W. Schultz, A.F. Taylor, and G.R. Smith. 1985. Chi-dependent DNA strand cleavage by RecBC enzyme. *Cell* 41:145–51.
Radding, C.M. 1991. Helical interactions in homologous pairing and strand exchange driven by RecA protein. *Journal of Biological Chemistry* 266:5355–58.
Radding, C.M., J. Flory, A. Wu, R. Kahn, C. DasGupta, D. Gonda, M. Bianchi, and S.S. Tsang. 1982. Three phases in homologous pairing: Polymerization of *recA* protein on single-stranded DNA, synapsis, and polar strand exchange. *Cold Spring Harbor Symposia on Quantitative Biology* 47:821–28.
Rafferty, J.B., S.E. Sedelnikova, D. Hargreaves, P.J. Artymiuk, P.J. Baker, G.J. Sharples, A.A. Mahdi, R.G. Lloyd, and D.W. Rice. 1996. Crystal structure of DNA recombination protein RuvA and a model for its binding to the Holliday junction. *Science* 274:415–21.
Roman, L. J., D.A. Dixon, and S.C. Kowalczykowski. 1991. RecBCD-dependent joint molecule formation promoted by the *Escherichia coli* RecA and SSB proteins. *Proceedings of the National Academy of Sciences USA* 88:3367–71.
Shah, R., R.J. Bennett, and S.C. West. 1994. Genetic recombination in *E. coli:* RuvC protein cleaves Holliday junctions at resolution hotspots in vitro. *Cell* 79:853–64.
Sun, H., D. Treco, N.P. Schultes, and J.W. Szostak. 1989. Double-strand breaks at an initiation site for meiotic gene conversion. *Nature* 338:87–90.
Yu, X., S.C. West, and E.H. Egelman. 1997. Structure and subunit composition of the RuvAB–Holliday junction complex. *Journal of Molecular Biology* 266:217–22.

第二十三章

General References and Reviews

Baltimore, D. 1985. Retroviruses and retrotransposons: The role of reverse transcription in shaping the eukaryotic genome. *Cell* 40:481–82.
Cohen, S.N. and J.A. Shapiro. 1980. Transposable genetic elements. *Scientific American* 242 (February):40–49.
Craig, N.L. 1996 V(D)J recombination and transposition: Closer than expected. *Science* 271:1512.
Doerling, H.-P. and P. Starlinger. 1984. Barbara McClintock's controlling elements: Now at theDNA level. *Cell* 39:253–59.
Eickbush, T.H. 2000. Introns gain ground. *Nature* 404:940–41.
Engels, W.R. 1983. The P family of transposable elements in *Drosophila*. *Annual Review of Genetics* 17:315–44.
Federoff, N.V. 1984. Transposable genetic elements in maize. *Scientific American* 250(June):84–99.
Grindley, N.G.F. and A.E. Leschziner. 1995. DNA transposition: From a black box to a color monitor. *Cell* 83:1063–66.
Kazazian, H.H., Jr. and J.V. Moran. 1998. The impact of L1 retrotransposons on the human genome. *Nature Genetics* 19:19–24.
Lambowitz, A.M. and S. Zimmerly. 2004. Mobile group II introns. *Annual Review of Genetics* 38:1–35.
Levin, K.L. 1997. It's prime time for reverse transcriptase. *Cell* 88:5–8.
Lewis, S.M. 1994. The mechanism of V(D)J joining: Lessons from molecular, immunological, and comparative analyses. *Advances in Immunology* 56:27–50.
Tonegawa, S. 1983. Somatic generation of antibody diversity. *Nature* 302:575–81.
Voytas, D.F. 1996. Retroelements in genome organization. *Science* 274:737–38.

Research Articles

Baltimore, D. 1970. Viral RNA-dependent DNA polymerase. *Nature* 226:1209–11.
Boland, S. and N. Kleckner, 1996. The three chemical steps of Tn10/IS10 transposition involve repeated utilization of a single active site. *Cell* 84:223–33.
Chapman, K.B., A.S. Byström, and J.D. Boeke. 1992. Initiator methionine tRNA is essential for Ty1 transcription. *Proceedings of the National Academy of Sciences USA* 89:3236–40.
Davies, D.R., I.Y. Goryshin, W.S. Reznikoff, and I. Rayment. 2000. Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science* 289:77–85.
Difilippantonio, M.J., C.J. McMahan, Q.M. Eastman, E. Spanopoulou, and D.G. Schatz. 1996. RAG1 mediates signal sequence recognition and recruitment of RAG2 in V(D)J recombination. *Cell* 87:253–62.
Garfinkel, D.J., J.F. Boeke, and G.R. Fink. 1985. Ty element transposition: Reverse transcription and virus-like particles. *Cell* 42:507–17.
Hesse, J.E., M.R. Lieber, K. Mizuuchi, and M. Gellert. 1989. V(D)J recombination: A functional definition of the joining signals. *Genes and Development* 3:1053–61.
Luan, D.D., M.H. Korman, J.L. Jakubczak, and T.H. Eickbush. 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: A mechanism for non-LTR retrotransposition. *Cell* 72:595–605.
Oettinger, M.A., D.G. Schatz, C. Gorka, and D. Baltimore. 1990. RAG-1 and RAG-2, adjacent genes that synergistically

activate V(D)J recombination. *Science* 248:1517–22.
Panet, A., W.A. Haseltine, D. Baltimore, G. Peters, F. Harada, and J. E. Dahlberg. 1975. Specific binding of tryptophan transfer RNA to avian myeloblastosis virus RNA-dependent DNA polymerase (reverse transcriptase). *Proceedings of the National Academy of Sciences USA* 72:2535–39.
Temin, H.M. and Mizutani, S. 1970. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* 226:1211–13.
Wessler, S.R. 1988. Phenotypic diversity mediated by the maize transposable elements *Ac* and *Spm*. *Science* 242:399–405.

第二十四章

General References and Reviews

Abbott, A. 1999. A post-genomic challenge: Learning to read patterns of protein synthesis. *Nature* 402:715–20.
Cheung, V.G., M. Morley, F. Aguilar, A. Massimi, R. Kucherlapati, and G. Childs. 1999. Making and reading microarrays. *Nature Genetics Supplement* 21:15–19.
Collins, F.S., M.S. Guyer, and A. Chakravarti. 1997. Variations on a theme: Cataloging human DNA sequence variation. *Science* 278:1580–81.
Goffeau, A. 1995. Life with 482 genes. *Science* 270:445–46.
Goffeau, A., B.G. Barrell, H. Bussey, R.W. Davis, B. Dujon, H. Feldmann, et al. 1996. Life with 6000 genes. *Science* 274:546–67.
Hieter, P. and Boguski, M. 1997. Functional genomics: It's all how you read it. *Science* 278:601–02.
Kumar, A. and M. Snyder. 2002. Protein complexes take the bait. *Nature* 415:123–24.
Lipshutz, R.J., S.P.A. Fodor, T.R. Gingeras, and D.J. Lockhart. 1999. High density synthetic oligonucleotide arrays. *Nature Genetics Supplement* 21:20–24.
Morell, V. 1996. Life's last domain. *Science* 273:1043–45.
Murray, T.H. 1991. Ethical issues in human genome research. *FASEB Journal* 5:55–60.
Reeves, R.H. 2000. Recounting a genetic story. *Nature* 405:283–34.
Service, R.F. 1998. Microchip arrays put DNA on the spot. *Science* 282:396–99.
Venter, J.C., H.O. Smith, and L. Hood. 1996. A new strategy for genome sequencing. *Nature* 381:364–66.
Young, R.A. 2000. Biomedical discovery with DNA arrays. *Cell* 102:9–15.
Zimmer, C. 2003. Tinker, tailor: Can Venter stitch together a genome from scratch? *Science* 299:1006–07.

Research Articles

Arbeitman, M.N., E.E.M. Furlong, F. Imam, E. Johnson, B.H. Null, B.S. Baker, M.A. Krasnow, M.P. Scott, R.W. Davis, and K.P. White. 2002. Gene expression during the life cycle of *Drosophila melanogaster. Science* 297:2270–75.
Blattner, F.R., G. Plunkett 3rd, C.A. Bloch, N.T. Perna, V. Burland, M. Riley, et al. 1997. The complete genomic sequence of *Escherichia coli* K12. *Science* 277:1453–62.
Bult, C.J., O. White, G.J. Olsen, L. Zhou, R.D. Fleischmann, G. G. Sutton, et al. 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii. Science* 273:1058–73.
Cheng, J., T.R. Gingeras, and many other authors. 2005. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* 308:1149–54.
C. elegans Sequencing Consortium, The. 1998. Genome sequence of the nematode *C. elegans:* A platform for investigating biology. *Science* 282:2013–18.
Deloukas, P., G.D. Schuler, G. Gyapay, E.M. Beasley, C. Soderlund, P. Rodriguez-Tome, et al. 1998. A physical map of 30,000 human genes. *Science* 282:744–46.
Dunah, A.W., H. Jeong, A. Griffin, Y.-M. Kim, D.G. Standaert, S.M. Hersch, M.M. Mouradian, A.B. Young, N. Tanese, and D. Krainc. 2002. Sp1 and TAFII130 transcriptional activity disrupted in early Huntington's disease. *Science* 296:2238–43.
Dunham, I., N. Shimizu, B.A. Roe, S. Chissoe, A.R. Hunt, J.E. Collins, (The Chromosome 22 Sequencing Consortium). 1999. The DNA sequence of human chromosome 22. *Nature* 402:489–95.
Gavin, A.-C., M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, et al. 2002. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415:141–47.
Glaever, G., A.M. Chu, L. Ni, C. Connelly, L. Riles, S. Veronneau, et al. 2002. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418:387–91.
Gusella, J.F., N.S. Wexler, P.M. Conneally, S.L. Naylor, M.A. Anderson, R.E. Tauzi, et al. 1983. A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* 306:234–38.
Hattori, M., A. Fujiyama, T.D. Taylor, H. Watanabe, T. Yada, H.S. Park, et al. (The Chromosome 21 Sequencing Consortium). 2000. The DNA sequence of human chromosome 21. *Nature* 405:311–19.
Ho, Y., A. Gruhler, A. Heilbut, G.D. Bader, L. Moore, S.L. Adams, et al. 2002. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 415:180–83.
Hudson, T.J., L.D. Stein, S.S. Gerety, J. Ma, A.B. Castle, J. Silva, et al. 1995. An STS-based map of the human genome. *Science* 270:1945–54.
Hutchinson, C.A. III, S.N. Peterson, S.R. Gill, R.T. Cline, O. White, C.M. Fraser, H.O. Smith, and J.C. Venter. 1999. Global transposon mutagenesis and a minimal mycoplasma genome. *Science* 286:2165–69.
International HapMap Consortium. 2005. A haplotype map of the human genome. *Nature* 437:1299–1320.
International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921.
Iyer, V.R., M.B. Eisen, D.T. Ross, G. Schuler, T. Moore, J. Lee, et al. 1999. The transcriptional program in the response to human fibroblasts to serum. *Science* 283:83–87.
Lim, L.P., N.C. Lau, P. Garrett-Engele, A. Grimson, J.M. Schelter, J. Castle, D.P. Bartel, P.S. Linsley, and J.M. Johnson. 2005. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* 433:769–73.
Mural, R.J., M.D. Adams, E.W. Myers, H.O. Smith, G.L. Miklos, R. Wides, et al. 2002. A comparison of whole-genome shotgun-derived mouse chromosome 16 and the human genome. *Science* 296:1661–71.
Ren, B., F. Robert, J.J. Wyrick, O. Aparicio, E.G. Jennings, I. Simon, et al. 2000. Genome-wide location and function of DNA binding proteins. *Science* 290:2306–09.
Ross, M.T. and many other authors. 2005. The DNA sequence of the human X chromosome. *Nature* 434:325–37.
Schuler, G.D., M.S. Boguski, E.A. Stewart, L.D. Stein, G. Gyapay, K. Rice, et al. 1996. A gene map of the human genome. *Science* 274:540–46.
Shizuya, H., B. Birren, U.-J. Kim, V. Mancino, T. Slepak, Y. Tachiiri, and M. Simon. 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proceedings of the National Academy of Sciences USA* 89:8794–97.
Sönnichsen, B. and many other authors. 2005. Full-genome RNAi profiling of early embryogenesis in *Caenorhabditis elegans. Nature* 434:462–69.
Stelzl, U., and many other authors. 2005. A human protein–protein interaction network: A resource for annotating the proteome. *Cell* 122:957–68.
Tong, A.H., B. Drees, G. Nardelli, G.D. Bader, B. Brannetti, L. Castagnoli, et al. 2002. A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. *Science* 295:321–4.

Velculescu, V.E., L. Zhang, B. Vogelstein, and K.W. Kinsler. 1995. Serial analysis of gene expression. *Science* 270:484–87.

Venter, J.C., M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, et al. 2001. The sequence of the human genome. *Science* 291:1304–51.

Wang, D.G., J.B. Fan, C.J. Sino, A. Berno, P. Young, R. Sapolsky, et al. 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280:1077–82.

Xie, X., J. Lu, E.J. Kulbokas, T.R. Golub, V. Mootha, K. Lindblad-Toh, E.S. Lander, and M. Kellis. 2005. Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. *Nature* 434:338–45.

Zhu, H., M. Bilgin, R. Bangham, D. Hall, A. Casamayor, P. Bertone, et al. 2001. Global analysis of protein activities using proteome chips. *Science* 293:2101–05.

索　引

A

AP 位点　605

B

BAC 步移　718
斑点膜　94
斑点印迹　94
半保留复制　6，38，579
半甲基化　44
胞质多聚腺苷酸化元件　444
报告探针　59
被加工的假基因　701
必需基因簇　727
表达序列标签　718
表达载体　60
表观修饰　462
表观遗传　338

C

cDNA 末端快速扩增法　55
Chargaff 法则　15
Charon 噬菌体　49
CpG 岛　706
C 值　23
C 值悖论　23
采取随机弥散型复制　579
插入序列　679
长散布因子　697
常染色体　3
常染色质　342
沉默突变　39
穿梭载体　64
纯合子　2

D

DNase 敏感区　327
DNase 足迹法　98
DNA 变性　20
DNA 分型　78
DNA 凝胶　245
DNA 熔解　20
DNA 微阵列　729
DNA 芯片　729
DNA 指纹　78
单倍体　2
单核苷酸多态性　742
蛋白质组学　704，728
等电点　72
等电聚焦　72
等位基因　1
等位基因特异性 RNAi　545
电泳阻滞分析　98
定点突变　51
定向克隆　49
端粒　645
端粒末端结合蛋白　649
端粒位置效应　342
短发夹 RNA　458
短干涉 RNA　451
短散布因子　701
多聚核糖体　401，566
多聚腺苷酸化　394
多聚腺苷酸化信号　405
多克隆位点　46

E

二倍体　2

F

发色底物　77
翻译　25
反密码子　572
反式剪接　437
反转录　691
反转录 PCR　56
反转录归巢　701
放射性自显影　74
非复制式转座　682
非模板链　25
非终止 mRNA　541
非转录间隔区　431
分子伴侣　179
辐射杂种作图法　718
复制起点　45
复制式转座　682

G

G 蛋白　535
杆状病毒　64
冈崎片段　582
隔离子　302
根癌农杆菌　52
功能基因组学　728
功能趋同进化　268
功能性 SELEX　101
共联体　682
共抑制　451
共转录剪切元件　425

寡核苷酸 53
冠瘿碱 65
冠瘿瘤 65
光密度计 75
滚环复制 588

H

核仁小 RNA 434
核糖核酸酶 H 53
核糖体循环因子 547
核小体 316
核小体定位 327
轰击法 64
琥珀密码子 536
琥珀突变 536
回文结构 43

J

Jun-Fos 二聚体 296
基因表达系列分析 733
基因间抑制 191
基因型 3
基因转换 676
基因组 704
基因组功能图谱 736
基因组文库 49
基因组学 704
基因座 3
加帽 394
假基因 440
碱基颠换 520
碱基转换 520
酵母人工染色体 52，714
结构基因组学 728
结构域 28
结构转录因子 269，300
截断转录 93
聚丙烯酰胺凝胶电泳 71
聚合酶链式反应 56
菌落杂交 55

K

Kozak 规则 484
抗终止 182
可变数串联重复序列 716
可读框 485
可诱导型表达载体 60
克隆 42
快速周转决定子 447

L

离子交换层析 73
连缀转录 94
联接体 657
两性电解质 72
亮氨酸拉链 288
裂解模式 181
磷屏成像仪 75
硫酸二甲酯 98
氯霉素乙酰转移酶 94
卵白石密码子 537
螺旋-环-螺旋 288

M

microRNA 463
每分钟的次数 76
每分钟蜕变量 75
孟德尔遗传 2
免疫沉淀法 97
免疫印迹 81
模板链 25
末端补平法 91
末端脱氧核苷酰转移酶 55
末端转移酶 55

N

Northern 印迹 89
N 蛋白利用位点 183
内部核糖体进入序列 487
黏粒 51
黏性末端 43
鸟枪法测序 718
凝胶电泳 69
凝胶过滤层析 73
凝胶阻滞分析 98

P

平末端 43

Q

qut 位点（Q 利用位点，Q utilization site） 187
前起始复合物 245
羟基自由基 99
羟基自由基足迹法 98
敲除小鼠 101
切口平移 53
亲和层析 73
亲和柱层析法 61
全保留复制 579
全染色体转录作图 733
缺失分析 736

R

R-M 系统 44
RNase 保护分析 92
RNase 作图 92
RNA 干涉 451
RNA 聚合酶 II 全酶 265
RNA 探针 92
RNA 诱导的沉默复合物 454
染色质 315
热循环仪 56
溶源模式 181

熔解温度 20
融合蛋白 60

S

S1 核酸酶 90
S1 图谱定位 90
Sanger 链末端终止测序法 81
SELEX 指数式富集的配体系统进化 100
Shine-Dalgarno 序列 478
SOS 响应 194
Southern 印迹 77
色谱（层析） 73
闪烁 76
闪烁液 76
上游激活因子 269
上游结合因子 269
生物信息学 704
十二烷基磺酸钠 71
实时定量 PCR 59
实时反转录 PCR 59
释放因子 538
噬菌斑 49，192
噬菌斑杂交 50
噬菌体 45
双链复制型 52
双脱氧核苷酸 81
双向凝胶电泳 72
羧基末端域 407

T

Taq 聚合酶 56
TATA 盒结合蛋白 247
t 环 650
替换载体 49
跳跃基因 682
铁应答元件 446
通用转录因子 245
同工 tRNA 517
同裂酶 43
同源结构域 287
同源重组 656
突变型 3
图谱 79
图位克隆技术 704

W

外切核酸酶 43
烷基化 602
微卫星序列 716
位点特异性重组 679
无 G 盒转录 93
无义介导的 mRNA 降解 544
无义相关的修饰剪接 544
物理图谱 83

X

洗脱 73
细菌人工染色体 52，715
限制性内切核酸酶 42
限制性片段长度多态性 79，705
限制性图谱 83
限制-修饰系统 44
硝化纤维素 97
小卫星 78
锌指 274，282
新裂酶 43
信使 RNA 25
性染色体 3
序列标签位点 716

Y

亚基变性 71
严谨性 52
摇摆假说 517
药物基因组学 742
野生型 3
液体闪烁计数 76
移码突变 514
遗传连锁 3
遗传密码 514
遗传作图 4
异裂酶 43
异染色质 342
易错旁路 615
引导 RNA 441
引物延伸 92
荧光原位杂交 80
萤光素酶 95
原病毒 692
原位表达分析 741
原位杂交 80

Z

杂合子 2
杂种败育 685
载体 45
增感屏 75
增强体 302
增强子 212
增色转换（效应） 20
赭石密码子 536
脂质体 64
纸层析 73
指形结构 283
质粒 45
重叠群 717
重组体 3
重组信号序列 688
重组修复 614
逐步克隆测序法 716

转化　64
转基因植株　66
转录　25
转录单位　403
转录间隔区　431
转录偶联核苷酸切除修复　608
转录停滞　267
转录暂停　267
转录组　728
转录组学　728
转染　64
转移信使 RNA　542
转座　679
转座酶　680
装配因子　269
着色性干皮病　608
自动测序仪　83
组蛋白密码　338
组分收集器　73
最小基因组　727

α 互补　48
3′-非翻译区　38
5′-非翻译区　38
β 钳　630